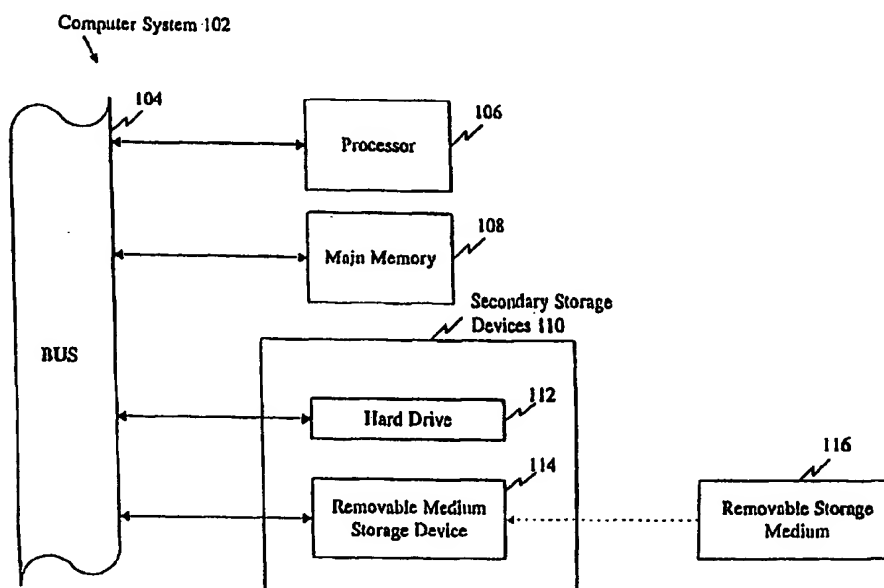




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/315, 16/12, C12Q 1/68		A2	(11) International Publication Number: WO 98/18931
			(43) International Publication Date: 7 May 1998 (07.05.98)
(21) International Application Number: PCT/US97/19588		(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).	
(22) International Filing Date: 30 October 1997 (30.10.97)			
(30) Priority Data: 60/029,960 31 October 1996 (31.10.96) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): KUNSCH, Charles, A. [US/US]; 2398B Dunwoody Crossing, Atlanta, GA 30338 (US). CHOI, Gil, H. [KR/US]; 11429 Potomac Oaks Drive, Rockville, MD 20850 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonville, MD 20882 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US). FANNON, Michael [US/US]; 13501 Rippling Brook Drive, Silver Spring, MD 20850 (US). DOUGHERTY, Brian, A. [US/US]; 708 Meadow Field Court, Mount Airy, MD 21771 (US).		Published Without international search report and to be republished upon receipt of that report.	

(54) Title: *STREPTOCOCCUS PNEUMONIAE* POLYNUCLEOTIDES AND SEQUENCES

(57) Abstract

The present invention provides polynucleotide sequences of the genome of *Streptococcus pneumoniae*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Streptococcus pneumoniae Polynucleotides and Sequences

FIELD OF THE INVENTION

5 The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Streptococcus pneumoniae*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation,
10 polypeptide production, assays and pharmaceutical development, among others.

BACKGROUND OF THE INVENTION

Streptococcus pneumoniae has been one of the most extensively studied
15 microorganisms since its first isolation in 1881. It was the object of many investigations that led to important scientific discoveries. In 1928, Griffith observed that when heat-killed encapsulated pneumococci and live strains constitutively lacking any capsule were concomitantly injected into mice, the nonencapsulated could be converted into encapsulated pneumococci with the same
20 capsular type as the heat-killed strain. Years later, the nature of this "transforming principle," or carrier of genetic information, was shown to be DNA. (Avery, O.T., et al., *J. Exp. Med.*, 79:137-157 (1944)).

 In spite of the vast number of publications on *S. pneumoniae* many questions about its virulence are still unanswered, and this pathogen remains a
25 major causative agent of serious human disease, especially community-acquired pneumonia. (Johnston, R.B., et al., *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991)). In addition, in developing countries, the pneumococcus is responsible for the death of a large number of children under the age of 5 years from pneumococcal pneumonia. The incidence of pneumococcal disease is highest in infants under 2
30 years of age and in people over 60 years of age. Pneumococci are the second most frequent cause (after *Haemophilus influenzae* type b) of bacterial meningitis and otitis media in children. With the recent introduction of conjugate vaccines for *H. influenzae* type b, pneumococcal meningitis is likely to become increasingly prominent. *S. pneumoniae* is the most important etiologic agent of community-

acquired pneumonia in adults and is the second most common cause of bacterial meningitis behind *Neisseria meningitidis*.

The antibiotic generally prescribed to treat *S. pneumoniae* is benzylpenicillin, although resistance to this and to other antibiotics is found occasionally. Pneumococcal resistance to penicillin results from mutations in its penicillin-binding proteins. In uncomplicated pneumococcal pneumonia caused by a sensitive strain, treatment with penicillin is usually successful unless started too late. Erythromycin or clindamycin can be used to treat pneumonia in patients hypersensitive to penicillin, but resistant strains to these drugs exist. Broad spectrum antibiotics (e.g., the tetracyclines) may also be effective, although tetracycline-resistant strains are not rare. In spite of the availability of antibiotics, the mortality of pneumococcal bacteremia in the last four decades has remained stable between 25 and 29%. (Gillespie, S.H., *et al.*, *J. Med. Microbiol.* 28:237-248 (1989).

S. pneumoniae is carried in the upper respiratory tract by many healthy individuals. It has been suggested that attachment of pneumococci is mediated by a disaccharide receptor on fibronectin, present on human pharyngeal epithelial cells. (Anderson, B.J., *et al.*, *J. Immunol.* 142:2464-2468 (1989). The mechanisms by which pneumococci translocate from the nasopharynx to the lung, thereby causing pneumonia, or migrate to the blood, giving rise to bacteremia or septicemia, are poorly understood. (Johnston, R.B., *et al.*, *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991).

Various proteins have been suggested to be involved in the pathogenicity of *S. pneumoniae*, however, only a few of them have actually been confirmed as virulence factors. Pneumococci produce an IgA1 protease that might interfere with host defense at mucosal surfaces. (Kornfield, S.J., *et al.*, *Rev. Inf. Dis.* 3:521-534 (1981). *S. pneumoniae* also produces neuraminidase, an enzyme that may facilitate attachment to epithelial cells by cleaving sialic acid from the host glycolipids and gangliosides. Partially purified neuraminidase was observed to induce meningitis-like symptoms in mice; however, the reliability of this finding has been questioned because the neuraminidase preparations used were probably contaminated with cell wall products. Other pneumococcal proteins besides neuraminidase are involved in the adhesion of pneumococci to epithelial and endothelial cells. These pneumococcal proteins have as yet not been identified. Recently, Cundell *et al.*, reported that peptide permeases can modulate

pneumococcal adherence to epithelial and endothelial cells. It was, however, unclear whether these permeases function directly as adhesions or whether they enhance adherence by modulating the expression of pneumococcal adhesions. (DeVelasco, E.A., *et al.*, *Micro. Rev.* 59:591-603 (1995). A better understanding of the virulence factors determining its pathogenicity will need to be developed to cope with the devastating effects of pneumococcal disease in humans.

Ironically, despite the prominent role of *S. pneumoniae* in the discovery of DNA, little is known about the molecular genetics of the organism. The *S. pneumoniae* genome consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory elements, such as prophages, plasmids, transposons and the like. Most physical characteristics and almost all of the genes of *S. pneumoniae* are unknown. Among the few that have been identified, most have not been physically mapped or characterized in detail. Only a few genes of this organism have been sequenced. (See, for instance current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. pneumoniae* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. pneumoniae*, infection involves the programmed expression of *S. pneumoniae* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. pneumoniae* genes and genomic organization would improve our understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. pneumoniae* would provide reagents for, among other things, detecting, characterizing and controlling *S. pneumoniae* infections. There is a need to characterize the genome of *S. pneumoniae* and for polynucleotides of this organism.

SUMMARY OF THE INVENTION

5 The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS:1-391.

10 The present invention provides the nucleotide sequence of several hundred contigs of the *Streptococcus pneumoniae* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-391.

15 The present invention further provides nucleotide sequences which are at least 95% identical to the nucleotide sequences of SEQ ID NOS:1-391.

20 The nucleotide sequence of SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-391 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

25 The present invention further provides systems, particularly computer-based systems which contain the sequence information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Streptococcus pneumoniae* genome.

30 Another embodiment of the present invention is directed to fragments of the *Streptococcus pneumoniae* genome having particular structural or functional attributes. Such fragments of the *Streptococcus pneumoniae* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs, fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression
35 modulating fragments or EMFs, and fragments which can be used to diagnose the

presence of *Streptococcus pneumoniae* in a sample, hereinafter referred to as diagnostic fragments or DFs.

Each of the ORFs in fragments of the *Streptococcus pneumoniae* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Streptococcus pneumoniae* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to isolated polypeptides and proteins encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention from cells which have been altered to express them.

The invention further provides methods of obtaining homologs of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as

a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both
5 monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples
10 derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

15 In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, or one of the DFs of the present invention; and
20 (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to
25 a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

30 The present genomic sequences of *Streptococcus pneumoniae* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Streptococcus pneumoniae* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Streptococcus pneumoniae* researchers

and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

DESCRIPTION OF THE FIGURES

FIGURE 1 is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

FIGURE 2 is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Streptococcus pneumoniae* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Streptococcus pneumoniae* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using Extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading

frames (ORFs) is accomplished by processing contigs with zorf or GenMark. The ORFs are searched against *S. pneumoniae* sequences from GenBank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990)). Results of the ORF
5 determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

10

The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-391. (As used herein, the "primary sequence" refers to the nucleotide
15 sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Streptococcus pneumoniae* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-391, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

20 As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-391" refers to any portion of the SEQ ID NOS:1-391 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Streptococcus pneumoniae* open reading frames (ORFs), expression modulating fragment (EMFs) and
25 fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in sample (DFs). A non-limiting identification of preferred representative fragments is provided in Tables 1-3. As discussed in detail below, the information provided in SEQ ID NOS:1-391 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled
30 in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Streptococcus pneumoniae* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-391 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances,
35 further investigation of a fragment or sequence of the invention may reveal a

nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-391. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-391 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-391 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-391 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining libraries and for sequencing are provided below, for instance. A wide variety of *Streptococcus pneumoniae* strains that can be used to prepare *S. pneumoniae* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC). While the present invention is enabled by the sequences and other information herein disclosed, the *S. pneumoniae* strain that provided the DNA of the present Sequence Listing, Strain 7/87 14.8.91, has been deposited in the ATCC, as a convenience to those of skill in the art. As a further convenience, a library of *S. pneumoniae* genomic DNA, derived from the same strain, also has been deposited in the ATCC. The *S. pneumoniae* strain was deposited on October 10, 1996, and was given Deposit No. 55840, and the cDNA library was deposited on October 11, 1996 and was given Deposit No. 97755. The genomic fragments in the library are 15 to 20 kb

fragments generated by partial Sau3A1 digestion and they are inserted into the BamHI site in the well-known lambda-derived vector lambda DASH II (Stratagene, La Jolla, CA). The provision of the deposits is not a waiver of any rights of the inventors or their assignees in the present subject matter.

5 The nucleotide sequences of the genomes from different strains of *Streptococcus pneumoniae* differ somewhat. However, the nucleotide sequences of the genomes of all *Streptococcus pneumoniae* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-391. Nearly all will be at least 99% identical and the great majority will be
10 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

15 Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988) can be used to generate the percent identity of nucleotide
20 sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-391, a representative
25 fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-391 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, provided refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.*,
30 a nucleotide sequence provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-391. Such a manufacture provides a large portion of the *Streptococcus pneumoniae* genome and parts thereof (*e.g.*, a *Streptococcus pneumoniae* open reading frame
35 (ORF)) in a form which allows a skilled artisan to examine the manufacture using

means not directly applicable to examining the *Streptococcus pneumoniae* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-

391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-391 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

5 The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Streptococcus pneumoniae* genome which contain homology to ORFs or proteins from both
10 *Streptococcus pneumoniae* and from other organisms. Among the ORFs discussed herein are protein encoding fragments of the *Streptococcus pneumoniae* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

 The present invention further provides systems, particularly computer-
15 based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Streptococcus pneumoniae* genome.

 As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence
20 information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention.

25 As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

 As used herein, "data storage means" refers to memory which can store
30 nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

 As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target
35 structural motif with the sequence information stored within the data storage

means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Streptococcus pneumoniae* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the

Streptococcus pneumoniae genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), is used to identify open reading frames within the *Streptococcus pneumoniae* genome. A skilled artisan can readily
5 recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of
10 embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114
15 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the
20 data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing
25 and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to isolated fragments of the *Streptococcus pneumoniae* genome. The fragments of the
5 *Streptococcus pneumoniae* genome of the present invention include, but are not limited to fragments which encode peptides and polypeptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in a
10 sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Streptococcus pneumoniae* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are
15 normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-391, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

20 A variety of purification means can be used to generate the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Streptococcus pneumoniae* DNA can be enzymatically sheared to produce fragments of 15-20 kb in length. These fragments can then be
25 used to generate a *Streptococcus pneumoniae* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS:1-391. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from
30 the lambda DNA library or *Streptococcus pneumoniae* genomic DNA. Thus, given the availability of SEQ ID NOS:1-391, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-391 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of
35 the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2, and 3 list ORFs in the *Streptococcus pneumoniae* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive, or more selective lists.

Table 1 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that over a continuous region of at least 50 bases are 95% or more identical (by BLAST analysis) to a nucleotide sequence available through GenBank in October, 1997.

Table 2 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through GenBank in October, 1997.

Table 3 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through GenBank in October, 1997.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the first nucleotide of the ORF (actually the first nucleotide of the stop codon immediately preceeding the ORF), counting from the 5' end of the contig strand; and the fourth column, "stop (nt)" indicates the last nucleotide of the stop codon defining the 3' end of the ORF.

In Tables 1 and 2, column five, lists the Reference for the closest matching sequence available through GenBank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column six in Tables 1 and 2 provides the gene name of the matching sequence; column seven provides the BLAST identity score and column eight the BLAST similarity score from the

comparison of the ORF and the homologous gene; and column nine indicates the length in nucleotides of the highest scoring segment pair identified by the BLAST identity analysis.

Each ORF described in the tables is defined by "start (nt)" (5') and "stop (nt)" (3') nucleotide position numbers. These position numbers refer to the boundaries of each ORF and provide orientation with respect to whether the forward or reverse strand is the coding strand and which reading frame the coding sequence is contained. The "start" position is the first nucleotide of the triplet encoding a stop codon just 5' to the ORF and the "stop" position is the last nucleotide of the triplet encoding the next in-frame stop codon (i.e., the stop codon at the 3' end of the ORF). Those of ordinary skill in the art appreciate that preferred fragments within each ORF described in the table include fragments of each ORF which include the entire sequence from the delineated "start" and "stop" positions excepting the first and last three nucleotides since these encode stop codons. Thus, polynucleotides set out as ORFs in the tables but lacking the three (3) 5' nucleotides and the three (3) 3' nucleotides are encompassed by the present invention. Those of skill also appreciate that particularly preferred are fragments within each ORF that are polynucleotide fragments comprising polypeptide coding sequence. As defined herein, "coding sequence" includes the fragment within an ORF beginning at the first in-frame ATG (triplet encoding methionine) and ending with the last nucleotide prior to the triplet encoding the 3' stop codon. Preferred are fragments comprising the entire coding sequence and fragments comprising the entire coding sequence, excepting the coding sequence for the N-terminal methionine. Those of skill appreciate that the N-terminal methionine is often removed during post-translational processing and that polynucleotides lacking the ATG can be used to facilitate production of N-terminal fusion proteins which may be beneficial in the production or use of genetically engineered proteins. Of course, due to the degeneracy of the genetic code many polynucleotides can encode a given polypeptide. Thus, the invention further includes polynucleotides comprising a nucleotide sequence encoding a polypeptide sequence itself encoded by the coding sequence within an ORF described in Tables 1-3 herein. Further, polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence to the foregoing polynucleotides, are contemplated by the present invention.

Polypeptides encoded by polynucleotides described above and elsewhere herein are also provided by the present invention as are polypeptide comprising a an amino acid sequence at least about 95%, preferably at least 97% and even more preferably 99% identical to the amino acid sequence of a polypeptide encoded by an
5 ORF shown in Tables 1-3. These polypeptides may or may not comprise an N-terminal methionine.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions 1, 3
10 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically
15 list percent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the percent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations
20 provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Streptococcus pneumoniae*
25 genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably
30 linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are
5 fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Streptococcus pneumoniae* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200
10 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Streptococcus pneumoniae* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a
15 target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or
20 a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below. A sequence which is suspected as being an EMF is
25 cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

30 As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Streptococcus pneumoniae* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Streptococcus pneumoniae* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers

consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most at least preferably 99.9% identical to SEQ ID NOS:1-391, with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated. Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Streptococcus pneumoniae* origin isolated by using part or all of the fragments in question as a probe or primer.

Preferred DFs of the present invention comprise at least about 17, preferably at least about 20, and more preferably at least about 50 contiguous nucleotides within an ORF set out in Tables 1-3. Most highly preferred DFs specifically hybridize to a polynucleotide containing the sequence of the ORF from which they are derived. Specific hybridization occurs even under stringent conditions defined elsewhere herein.

Each of the ORFs of the *Streptococcus pneumoniae* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particularly *Streptococcus pneumoniae*. Especially preferred in this regard are ORFs such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Streptococcus pneumoniae*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Streptococcus pneumoniae*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56:560 (1991) and *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK, pBS KS, pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG

(available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers.

5 Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

10 The present invention further provides host cells containing any one of the isolated fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or
15 a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation,
20 which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated
25 fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF. The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ
30 from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs and subfragments thereof depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

10 "Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Streptococcus pneumoniae* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

15 Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancer and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

25 "Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express

heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas* and *Streptomyces*. Others may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication

derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Streptococcus pneumoniae*, of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Streptococcus pneumoniae* is defined as a homolog of a fragment of the *Streptococcus pneumoniae* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Streptococcus pneumoniae* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-391 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ

ID NOS:1-391 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.*,
5 *PCR Protocols*, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-391 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50-
10 65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

15 When using DNA probes derived from SEQ ID NOS:1-391, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, hybridizing at 50- 65°C in 5X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences
20 having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

25 Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Streptococcus pneumoniae*.

30 ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and
35 industrial purposes consistent with the type of putative identification of the

polypeptide. Such identifications permit one skilled in the art to use the *Streptococcus pneumoniae* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., MacMillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESSES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Streptococcus pneumoniae* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-391.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*,

Symbiosis 21:79 (1986) and Voragen *et al.* in *Biocatalysts In Agricultural Biotechnology*, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series* 389:93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Streptococcus pneumoniae*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1:21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in *GENE MANIPULATIONS AND FUNGI*; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta.* 872:83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteases Structure Function and*
5 *Biology*, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes by 1990*, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical*
10 *Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of
15 lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates.
20 Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of
25 interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

30 When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other

hand, has been described in detail by Bud *et al.*, Chemistry in Britain (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination.

2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques In Biochemistry And Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* 4:72 (1983), pgs. 77-96 of Cole *et al.*, in *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc. (1985)). Any animal (mouse, rabbit, etc.) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods

include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

5 The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

10 For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/O-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., 20 *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of 25 the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above-described antibodies in 30 detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labeling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308 35

(1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W., *J. Immunol. Meth.* 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays to identify cells or tissues in which a fragment of the
5 *Streptococcus pneumoniae* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for
10 coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays as well as for
15 immunoaffinity purification of the proteins of the present invention.

3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample,
20 using one of the DFs or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the DFs of the present invention and assaying for binding of the DFs or antibodies to components within the test sample.

Conditions for incubating a DF or antibody with a test sample vary.
25 Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs or antibodies of the present invention. Examples of such assays
30 can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and*

Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

5 The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

10 In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the DFs or antibodies of the present invention; and (b) one or
15 more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to
20 efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay,
25 containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, *etc.*), and containers which contain the reagents used to detect the bound antibody or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the
30 enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed DFs and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Streptococcus pneumoniae* fragment and contigs herein
5 described.

In general, such methods comprise steps of:

- (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Streptococcus pneumoniae* genome; and
- 10 (b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

- 15 For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is
20 chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides," in
25 *Synthetic Peptides, A User's Guide*, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, *Biochemistry* 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one
30 of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

5 Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem.* 56:560 (1991);
10 *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the
15 sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be
20 used to modulate the growth or pathogenicity of *Streptococcus pneumoniae*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents
25 which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth pathogenicity of *Streptococcus pneumoniae* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of
30 the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth by binding to an important protein thus blocking the biological
35 activity of the protein, while other agents may bind to a component of the outer

surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of a vaccine based on outer membrane components are well known in the art.

As used herein, a "related organism" is a broad term which refers to any organism whose growth can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may

be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s)

serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in a mixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, *e.g.*, human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-

microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES
5 (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or
10 sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

15 6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing
20 protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present
25 disclosure.

ILLUSTRATIVE EXAMPLES

LIBRARIES AND SEQUENCING

30 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2:231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability, P , that any given base in a sequence of size L , in
35 nucleotides, is not sequenced after a certain amount, n , in nucleotides, of random

sequence has been determined can be calculated by the equation $P = e^{-m}$, where m is L/n , the fold coverage. For instance, for a genome of 2.8 Mb, $m=1$ when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point, $P = e^{-1} = 0.37$. The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence L has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size L , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length, G , is determined by the equation $G = Le^{-m}$, and the average gap size, g , follows the equation, $g = L/n$. Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

Streptococcus pneumoniae DNA is prepared by phenol extraction. A mixture containing 200 µg DNA in 1.0 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 50% glycerol is processed through a nebulizer (IPI Medical Products) with a stream of nitrogen adjusted to 35 Kpa for 2 minutes. The sonicated DNA is ethanol precipitated and redissolved in 500 µl TE buffer.

To create blunt-ends, a 100 µl aliquot of the resuspended DNA is digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 µl BAL31 buffer. The digested DNA is phenol-extracted, ethanol-precipitated, redissolved in 100 µl TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size is excised from the gel, and the LGT agarose is melted and the resulting solution is extracted with phenol to separate the agarose from the

DNA. DNA is ethanol precipitated and redissolved in 20 μ l of TE buffer for ligation to vector.

A two-step ligation procedure is used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 μ l) contains 2 μ g of DNA fragments, 2 μ g pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and is incubated at 14°C for 4 hr. The ligation mixture then is phenol extracted and ethanol precipitated, and the precipitated DNA is dissolved in 20 μ l TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder are visualized by ethidium bromide-staining and UV illumination and identified by size as insert (I), vector (v), v+I, v+2i, v+3i, etc. The portion of the gel containing v+I DNA is excised and the v+I DNA is recovered and resuspended into 20 μ l TE. The v+I DNA then is blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50 μ l) containing the v+I linears, 500 μ M each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+I linears are dissolved in 20 μ l TE. The final ligation to produce circles is carried out in a 50 μ l reaction containing 5 μ l of v+I linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture is stored at -20°C.

This two-stage procedure results in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies 3 (1):5* (1990)) are used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells are plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating is carried out as follows. A 100 μ l aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) is thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 μ l aliquot of 1.42 M beta-mercaptoethanol is added to the aliquot of cells to a final concentration of 25 mM. Cells are incubated on ice for 10 min. A 1 μ l aliquot of the final ligation is added to the cells and incubated on ice for 30 min. The cells are heat pulsed for 30 sec. at 42°C and

placed back on ice for 2 min. The outgrowth period in liquid culture is eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture is plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl₂ (1 M), and 1 ml MgSO₄ /100 ml SOB agar. The 15 ml top layer is poured just prior to plating. Our titer is approximately 100 colonies/10 µl aliquot of transformation.²⁴

All colonies are picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products are deleted from the library, resulting in a slight increase in gap number over that expected.

3. Random DNA Sequencing

High quality double stranded DNA plasmid templates are prepared using a "boiling bead" method developed in collaboration with Advanced Genetic Technology Corp. (Gaithersburg, MD) (Adams *et al.*, *Science* 252:1651 (1991); Adams *et al.*, *Nature* 355:632 (1992)). Plasmid preparation is performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Template concentration is determined using Hoechst Dye and a Millipore Cytofluor. DNA concentrations are not adjusted, but low-yielding templates are identified where possible and not sequenced.

Templates are also prepared from two *Streptococcus pneumoniae* lambda genomic libraries. An amplified library is constructed in the vector Lambda GEM-12 (Promega) and an unamplified library is constructed in Lambda DASH II (Stratagene). In particular, for the unamplified lambda library, *Streptococcus pneumoniae* DNA (> 100 kb) is partially digested in a reaction mixture (200 µl) containing 50 µg DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23°C. The digested DNA was phenol-extracted and electrophoresed on a 0.5% low melting agarose gel at 2V/cm for 7 hours. Fragments from 15 to 25 kb are excised and recovered in a final volume of 6 µl. One µl of fragments is used with 1 µl of DASHII vector (Stratagene) in the recommended ligation reaction. One µl of the ligation mixture is used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract (Stratagene, #227711). Phage

are plated directly without amplification from the packaging mixture (after dilution with 500 μ l of recommended SM buffer and chloroform treatment). Yield is about 2.5×10^3 pfu/ μ l. The amplified library is prepared essentially as above except the lambda GEM-12 vector is used. After packaging, about 3.5×10^4 pfu are plated on the restrictive NM539 host. The lysate is harvested in 2 ml of SM buffer and stored frozen in 7% dimethylsulfoxide. The phage titer is approximately 1×10^9 pfu/ml.

Liquid lysates (100 μ l) are prepared from randomly selected plaques (from the unamplified library) and template is prepared by long-range PCR using T7 and T3 vector-specific primers.

Sequencing reactions are carried out on plasmid and/or PCR templates using the AB Catalyst LabStation with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers (Adams *et al.*, *Nature* 368:474 (1994)). Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. T7 and SP6 primers are used to sequence the ends of the inserts from the Lambda GEM-12 library and T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are performed by eight individuals using an average of fourteen AB 373 DNA Sequencers per day. All sequencing reactions are analyzed using the Stretch modification of the AB 373, primarily using a 34 cm well-to-read distance. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

Richards *et al.*, Chapter 28 in AUTOMATED DNA SEQUENCING AND ANALYSIS, M. D. Adams, C. Fields, J. C. Venter, Eds., Academic Press, London, (1994) described the value of using sequence from both ends of sequencing templates to facilitate ordering of contigs in shotgun assembly projects of lambda and cosmid clones. We balance the desirability of both-end sequencing (including the reduced cost of lower total number of templates) against shorter read-lengths for sequencing reactions performed with the M13RP1 (reverse) primer compared to the M13-21 (forward) primer. Approximately one-half of the templates are sequenced from both ends. Random reverse sequencing reactions are

done based on successful forward sequencing reactions. Some M13RP1 sequences are obtained in a semi-directed fashion: M13-21: sequences pointing outward at the ends of contigs are chosen for M13RP1 sequencing in an effort to specifically order contigs.

5

4. Protocol for Automated Cycle Sequencing

The sequencing is carried out using ABI Catalyst robots and AB 373 Automated DNA Sequencers. The Catalyst robot is a publicly available sophisticated pipetting and temperature control robot which has been developed specifically for DNA sequencing reactions. The Catalyst combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates are combined in the wells of an aluminum 96-well thermocycling plate. Thirty consecutive cycles of linear amplification (*i.e.*, one primer synthesis) steps are performed including denaturation, annealing of primer and template, and extension; *i.e.*, DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols are used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer is labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 DNA Sequencer for electrophoresis, detection, and base-calling. ABI currently supplies pre-mixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions are loaded per AB373 Sequencer each day, for a total of 960 samples. Electrophoresis is run overnight following the manufacturer's protocols, and the data is collected for twelve hours. Following electrophoresis and fluorescence detection, the ABI 373 performs automatic lane tracking and base-calling. The lane-tracking is confirmed visually. Each sequence electropherogram (or fluorescence lane trace) is inspected visually and assessed for quality. Trailing

sequences of low quality are removed and the sequence itself is loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence is removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 are around 400 bp and depend mostly on the quality of the template used for the sequencing reaction. ABI 373 Sequencers converted to Stretch Liners provide a longer electrophoresis path prior to fluorescence detection and increase the average number of usable bases to 500-600 bp.

INFORMATICS

1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi- user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 10^4 fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching

fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164:765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

The process resulted in 391 contigs as represented by SEQ ID NOs:1-391.

3. Identifying Genes

The predicted coding regions of the *Streptococcus pneumoniae* genome were initially defined with the program GeneMark, which finds ORFs using a probabilistic classification technique. The predicted coding region sequences were used in searches against a database of all nucleotide sequences from GenBank (October, 1997), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases. ORFs that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

ILLUSTRATIVE APPLICATIONS

1. Production of an Antibody to a *Streptococcus pneumoniae* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.*, *Basic Methods in Molecular Biology*, Elsevier, New York. Section 21-2 (1989).

3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, antibodies are useful in various animal models of pneumococcal disease as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic or immunoprophylactic reagent.

4. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Streptococcus pneumoniae* genome, such as those of Tables 1-3 and SEQ ID NOS:1-391 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

10

5. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Streptococcus pneumoniae* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

15
20

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Streptococcus pneumoniae* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Streptococcus pneumoniae* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Streptococcus pneumoniae* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Streptococcus pneumoniae* DNA 3' primer, taking care to ensure that the *Streptococcus pneumoniae* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Streptococcus pneumoniae* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Streptococcus pneumoniae* DNA.

Alternatively and if antibody production is not possible, the *Streptococcus pneumoniae* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease
5 cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Streptococcus pneumoniae* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed
10 transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or
15 Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* Express™ Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of
20 the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1	1	437	1003	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	92	200	567
2	5	6169	5720	gb U04047	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence ISI202 transposase gene, complete cds	96	450	450
2	6	6592	6167	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	98	426	426
3	11	9770	9147	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	624	624
3	12	10489	9671	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	91	819	819
3	13	11346	12019	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	474	474
3	14	12017	13375	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1359	1359
3	15	13421	14338	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	918	918
3	16	14329	15171	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	843	843
3	17	15132	17282	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	2151	2151
3	18	17267	18397	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1069	1131
4	1	46	1188	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORP3 and ORF5	99	1143	1143
4	2	1198	2529	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORP3 and ORF5	99	876	1332
5	7	11297	11473	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	82	175	177
6	7	7125	7364	emb Z77726 SPIS	S.pneumoniae DNA for insertion sequence ISI318 (1372 bp)	93	238	240
6	8	7322	7570	emb Z77725 SPIS	S.pneumoniae DNA for insertion sequence ISI381 (966 bp)	95	160	249
6	9	7533	7985	emb Z77725 SPIS	S.pneumoniae DNA for insertion sequence ISI381 (966 bp)	99	453	453
6	123	20197	19733	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	96	465	465
7	10	8305	7682	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	95	624	624

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
7	11	9024	8206	emb 283335 SP28	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	95	819	819
10	13	9304	8078	gb L29323	Streptococcus pneumoniae methyl transferase (mtr) gene cluster, complete cds	93	513	1227
11	2	548	919	emb 279691 SOOR	S.pneumoniae yorI(A,B,C,D,E), ftstL, pbpX and regR genes	99	316	372
11	3	892	1980	emb 279691 SOOR	S.pneumoniae yorI(A,B,C,D,E), ftstL, pbpX and regR genes	99	1089	1089
11	5	3040	3477	emb 279691 SOOR	S.pneumoniae yorI(A,B,C,D,E), ftstL, pbpX and regR genes	99	259	438
11	6	3480	3247	emb 279691 SOOR	S.pneumoniae yorI(A,B,C,D,E), ftstL, pbpX and regR genes	99	234	234
11	7	3601	4557	emb 279691 SOOR	S.pneumoniae yorI(A,B,C,D,E), ftstL, pbpX and regR genes	98	957	957
11	8	4506	4886	emb 279691 SOOR	S.pneumoniae yorI(A,B,C,D,E), ftstL, pbpX and regR genes	99	381	381
11	9	4884	7142	emb X16367 SPPB	Streptococcus pneumoniae pbpX gene for penicillin binding protein 2X	99	2259	2259
11	10	7132	8124	emb X16367 SPPB	Streptococcus pneumoniae pbpX gene for penicillin binding protein 2X	98	70	993
13	1	53	1126	gb W31296	S.pneumoniae recP gene, complete cds	99	437	1074
14	3	1837	2148	emb 283335 SP28	S.pneumoniae dexB, capI(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	87	96	312
14	4	2518	2108	gb W36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	411	411
15	9	8942	8511	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19fABCDCEFGHIJKLMNO) genes, complete cds, and aliA gene, partial cds	89	340	432
17	7	3910	3458	emb 277726 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	453	453
17	8	4304	3873	emb 277727 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	96	382	432
19	1	41	529	emb X94909 SPIG	S.pneumoniae iga gene	75	368	489
19	2	554	757	gb L07752	Streptococcus pneumoniae attachment site (attB), DNA sequence	99	167	204
19	3	946	1827	gb L07752	Streptococcus pneumoniae attachment site (attB), DNA sequence	94	100	882
20	1	937	182	gb U33315	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor (comC), histidine protein kinase (comD) and response regulator (comE) genes, complete cds, tRNA-Arg and tRNA-Gln genes	99	756	756
20	2	2271	931	gb U33315	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor (comC), histidine protein kinase (comD) and response regulator (comE) genes, complete cds, tRNA-Arg and tRNA-Gln genes	98	1341	1341

S. pneumoniae - Coding regions containing known sequences

TABLE 1

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
20	3	3175	2684	gb U76218	Streptococcus pneumoniae competence stimulating peptide precursor ComC (comC), histidine kinase homolog ComD (comD), and response regulator homolog ComE (comE) genes, complete cds	99	492	492
20	4	3322	4527	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtre), SPSP0J (spsp0j), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	1206	1206
20	5	4573	5343	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtre), SPSP0J (spsp0j), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	771	771
20	6	5532	6917	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtre), SPSP0J (spsp0j), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	1386	1386
20	7	5995	8212	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtre), SPSP0J (spsp0j), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	1218	1218
20	8	8214	8471	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtre), SPSP0J (spsp0j), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	98	258	258
20	9	8534	9670	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtre), SPSP0J (spsp0j), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnan) genes, complete cds	99	1137	1137
22	14	11887	12267	emb Z77726 SPIS	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	99	226	381
22	15	12708	12256	emb Z77727 SPIS	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	453
22	16	13165	12662	emb Z77726 SPIS	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	504	504
22	23	18398	18910	emb Z86112 SPZ8	S. pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	95	463	513
22	24	18829	19299	emb Z86112 SPZ8	S. pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	99	443	471
23	5	5624	4203	emb X52474 SPPL	S. pneumoniae ply gene for pneumolysin	99	1422	1422
23	6	6063	5629	gb M17717	S. pneumoniae pneumolysin gene, complete cds	98	197	435
26	1	5500	2	emb X94909 SPIG	S. pneumoniae iga gene	87	3487	5499
26	2	5823	5584	gb U47687	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	99	151	240
26	3	6878	5685	gb U47687	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	100	50	1194

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
26	8	14698	14854	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	99	338	357
26	9	14763	14924	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	100	94	162
26	10	14922	15173	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
28	1	80	505	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	99	426	426
28	2	503	952	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	450	450
28	3	780	1298	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	181	519
34	1	207	1523	gb U08611	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	99	1317	1317
34	2	1477	2367	gb U08611	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	96	795	891
34	3	2593	3420	gb U21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	446	828
34	4	2790	2647	gb U21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	98	137	144
34	5	3418	4416	gb U21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	999	999
34	9	7764	7507	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	93	201	258
34	16	10562	10257	emb X63602 SP80	S.pneumoniae mmsA-Box	92	238	306
35	4	1176	1439	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	87	248	264
35	5	1458	1961	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19A,B,C,D,E,F,G,H,I,J,K,L,M,N,O) genes, complete cds, and alia gene, partial cds	98	264	504
35	17	16172	15477	emb X85787 SPCP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, cps14M, cps14N, cps14O genes	97	696	696
35	18	16961	16170	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	86	792	792
35	19	17620	16871	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19A,B,C,D,E,F,G,H,I,J,K,L,M,N,O) genes, complete cds, and alia gene, partial cds	83	750	750

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
35	20	19061	17604	emb X85787 SPCP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, cps14M, cps14N genes	94	1458	1458
36	19	18960	18352	gb U40786	Streptococcus pneumoniae surface antigen A variant precursor (psaA) and 18 kDa protein genes, complete cds, and ORF1 gene, partial cds	99	609	609
36	20	19914	18966	gb U53509	Streptococcus pneumoniae surface adhesin A precursor (psaA) gene, complete cds	99	969	969
37	1	2743	179	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	2565	2565
37	2	2985	2824	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	162	162
37	3	5034	3070	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	1965	1965
37	4	5134	5790	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	657	657
37	5	6171	5813	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	96	339	339
38	19	12969	13268	gb M28679	S.pneumoniae promoter region DNA	100	64	300
39	2	1256	2137	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	882	882
39	3	2405	3370	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	966	966
40	9	5253	7208	gb M29686	S.pneumoniae mismatch repair (hexB) gene, complete cds	99	1956	1956
41	1	3	1037	emb 217307 SPRE	S.pneumoniae recA gene encoding RecA	99	1027	1035
41	2	1328	2713	emb 234303 SPCI	Streptococcus pneumoniae cin operon encoding the cinA, recA, dinF, lytA genes, and downstream sequences	99	1386	1386
41	3	3083	4045	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	99	963	963
41	4	3272	3096	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	100	177	177
41	5	3603	3860	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	100	258	258
41	6	4755	5162	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	408	408
41	7	5270	5716	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	447	447
41	8	6112	6918	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	431	807
41	9	6916	7119	gb L36660	Streptococcus pneumoniae ORF, complete cds	100	204	204
41	10	7082	7660	gb L36660	Streptococcus pneumoniae ORF, complete cds	97	552	579
41	11	7680	7979	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	81	300
41	12	9169	8717	emb 277727 SPIS	S.pneumoniae DNA for insertion sequence IS118 (823 bp)	97	353	453

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
41	13	9533	9132	emb 277725 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
41	14	9669	9475	emb 282001 SP28	S.pneumoniae pcps gene and open reading frames	100	189	195
44	5	7190	7555	emb 282001 SP28	S.pneumoniae pcps gene and open reading frames	99	366	366
44	6	8059	7607	emb 277726 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	97	453	453
44	7	8423	8022	emb 277725 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
44	8	8559	8365	emb 282001 SP28	S.pneumoniae pcps gene and open reading frames	100	189	195
48	9	6480	4687	gb L39074	Streptococcus pneumoniae pyruvate oxidase (spxB) gene, complete cds	99	1794	1794
49	2	231	2603	gb L20561	Streptococcus pneumoniae Exp7 gene, partial cds	100	216	2373
53	6	2407	2156	gb U04047	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
53	7	2566	2405	emb 283335 SP28	S.pneumoniae dexB, cap A,B,C,D,E,F,G,H,I,J,K genes, dTDP-rhamnose biosynthesis genes and aliA gene	100	94	162
53	8	2831	2475	emb 283335 SP28	S.pneumoniae dexB, cap A,B,C,D,E,F,G,H,I,J,K genes, dTDP-rhamnose biosynthesis genes and aliA gene	99	338	357
54	13	12409	11105	emb 283335 SP28	S.pneumoniae dexB, cap A,B,C,D,E,F,G,H,I,J,K genes, dTDP-rhamnose biosynthesis genes and aliA gene	67	591	1305
55	22	120488	19949	emb 284379 HS28	S.pneumoniae dfr gene (isolate 92)	99	540	540
61	11	11864	9900	emb 216082 PNAI	Streptococcus pneumoniae aliB gene	98	1965	1965
63	1	3	239	gb M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	100	237	237
63	2	233	2611	gb M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	99	2330	2379
63	3	2557	2823	gb M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	99	266	267
63	4	2958	4664	gb M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	95	69	1707
67	6	3770	3399	gb L20670	Streptococcus pneumoniae hyaluronidase gene, complete cds	96	372	372
67	7	7161	4171	gb L20670	Streptococcus pneumoniae hyaluronidase gene, complete cds	99	2938	2991
70	1	1	702	gb M14340	S.pneumoniae DpnI gene region encoding dpcN and dpcD, complete cds	100	693	702
70	2	678	1160	gb M14340	S.pneumoniae DpnI gene region encoding dpcN and dpcD, complete cds	100	483	483
70	3	2490	1210	gb M14339	S.pneumoniae DpnII gene region encoding dpcM, dpcA, dpcB, complete cds	98	462	1281
70	7	4230	4424	gb J04234	S.pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	147	195
70	8	5197	4316	gb J04234	S.pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	881	882

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
70	13	8108	9874	gb L20562	Streptococcus pneumoniae Exo8 gene, partial cds	93	234	1767
71	22	27964	28341	emb X63602 SP80	S.pneumoniae mmsA-Box	93	233	378
72	5	4607	3452	emb Z26850 SPAT	S.pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	97	102	1056
73	1	471	113	emb X63602 SP80	S.pneumoniae mmsA-Box	91	193	339
73	3	3658	977	gb J04479	S.pneumoniae DNA polymerase I (polA) gene, complete cds	99	2682	2682
73	8	4864	5379	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	318	516
77	3	2622	1999	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	95	624	624
77	4	3341	2523	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	91	819	819
78	1	341	3	emb X77249 SPR6	S.pneumoniae (R6) clbA/ciaH genes	99	339	339
78	2	1095	325	emb X77249 SPR6	S.pneumoniae (R6) clbA/ciaH genes	99	771	771
82	10	11436	10816	gb U90721	Streptococcus pneumoniae signal peptidase I (spi) gene, complete cds	97	621	621
82	11	12402	11434	gb U93576	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	98	953	969
82	12	12381	12704	gb U93576	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	100	51	324
83	8	3212	3550	emb Z7727 SPIS	S.pneumoniae DNA for insertion sequence ISI318 (823 bp)	97	290	339
83	10	4662	6851	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	2190	2190
83	11	6849	8213	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	1365	1365
83	12	8236	9090	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	855	855
83	13	9283	13017	gb L15190	Streptococcus pneumoniae SAICAR synthetase (purC) gene, complete cds	100	107	3735
83	23	22147	23313	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (atrII) gene, complete cds	98	218	1167
83	24	23268	23450	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (atrII) gene, complete cds	98	172	183
83	25	27527	23505	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (atrII) gene, complete cds	99	3826	4023

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
83	26	28472	27771	gb U36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (strH) gene, complete cds	99	416	702
84	4	4554	6173	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	98	697	1620
87	6	5951	5316	emb Z77725 SP15	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	439	636
88	5	2957	3511	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	555	555
88	6	3466	4269	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) aqo SAICAR synthetase (purC) genes, complete cds	94	804	804
89	13	9878	10093	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	211	216
89	14	10062	10412	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	97	335	351
93	10	5303	4941	emb X63602 SP80	S.pneumoniae mmsA-Box	89	237	363
97	4	1708	1520	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	91	140	189
99	1	89	700	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	93	592	612
99	2	1773	775	emb X17337 SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	99	998	999
99	3	2794	1712	emb X17337 SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	99	1083	1083
99	4	3732	2788	emb X17337 SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	100	945	945
99	5	5249	3714	emb X17337 SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	100	1536	1536
99	6	7262	5277	emb X17337 SPAM	Streptococcus pneumoniae ami locus conferring aminopterin resistance	99	1986	1986
101	1	216	1538	emb X54225 SPEN	S.pneumoniae epuA and endA genes for 7 kda protein and membrane endonuclease	99	146	1323
101	2	1492	1719	emb X54225 SPEN	S.pneumoniae epuA and endA genes for 7 kda protein and membrane endonuclease	99	228	228
101	3	1694	1855	emb X54225 SPEN	S.pneumoniae epuA and endA genes for 7 kda protein and membrane endonuclease	100	162	162
101	4	1701	2582	emb X54225 SPEN	S.pneumoniae epuA and endA genes for 7 kda protein and membrane endonuclease	100	882	882
103	7	5556	5041	emb Z95914 SP29	Streptococcus pneumoniae soda gene	100	396	516
104	2	1347	1556	emb Z77727 SP15	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	83	206	210

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
105	5	5381	5028	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	98	353	354
105	6	6089	5379	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	98	84	711
107	4	2785	1880	emb X16022 SPPE	S.pneumoniae penA gene	98	72	906
107	5	2913	4988	emb X16022 SPPE	S.pneumoniae penA gene	99	1692	2076
107	6	4981	5595	emb X13136 SPPE	Streptococcus pneumoniae penA gene for penicillin binding protein 2B lacking N-term. (penicillin resistant strain)	91	107	615
108	9	9068	8718	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	342	351
108	12	11308	10922	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	199	387
109	3	2768	2241	emb 277725 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	61	528
109	4	2688	2855	emb 277726 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	96	148	168
109	5	2862	3269	emb 277727 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	408
109	6	5320	3584	gb M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	100	371	1737
113	1	431	3	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	429	429
113	10	9788	8532	emb X99400 SPDA	S.pneumoniae dacA gene and ORF	99	1257	1257
113	11	9870	10985	emb X99400 SPDA	S.pneumoniae dacA gene and ORF	99	1116	1116
114	3	2530	2030	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	481	501
115	11	11303	10932	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	372	372
117	1	897	3302	emb X72967 SPNA	S.pneumoniae nanA gene	99	2402	2406
117	2	3277	3831	emb X72967 SPNA	S.pneumoniae nanA gene	99	237	555
117	3	4327	3899	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	429	429
121	2	1369	1941	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	202	573
121	3	2412	4253	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	1842	1842
122	8	5066	5587	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	64	451	522

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
125	1	1811	189	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	92	99	1623
128	15	12496	11204	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	91	705	1293
134	1	1	492	emb Y10818 SPV1	S.pneumoniae spsA gene	99	203	492
134	2	556	2652	gb AF019904	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	86	685	2097
134	3	1160	837	emb Y10818 SPV1	S.pneumoniae spsA gene	86	324	324
134	4	3952	2882	gb AF019904	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	98	215	1071
134	8	7992	9848	gb U12567	Streptococcus pneumoniae P13 glycerol-3-phosphate dehydrogenase (gipD) gene, partial cds, and glycerol uptake facilitator (gipf) and ORF3 genes, complete cds	99	285	1857
134	9	9846	10622	gb U12567	Streptococcus pneumoniae P13 glycerol-3-phosphate dehydrogenase (gipD) gene, partial cds, and glycerol uptake facilitator (gipf) and ORF3 genes, complete cds	99	570	777
134	10	10805	11122	gb U12567	Streptococcus pneumoniae P13 glycerol-3-phosphate dehydrogenase (gipD) gene, partial cds, and glycerol uptake facilitator (gipf) and ORF3 genes, complete cds	100	318	318
137	13	7970	8443	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cpsA)(ABCDEFHJKLMO) genes, complete cds, and alia gene, partial cds	90	420	474
137	14	8590	8775	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	174	186
137	15	8773	8967	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	98	195	195
137	16	9223	9687	emb 277726 SP15	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	96	446	465
137	17	9641	10051	emb 277727 SP15	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	96	293	411
139	10	12998	12702	emb X63602 SP80	S.pneumoniae mmsA-Box	90	234	297
141	8	7805	8938	emb Z49988 SPM	Streptococcus pneumoniae mmsA gene	99	338	1114
141	9	8936	10972	emb Z49988 SPM	Streptococcus pneumoniae mmsA gene	99	2037	2037
141	10	11472	12467	emb Z49988 SPM	Streptococcus pneumoniae mmsA gene	100	76	996
142	2	257	814	gb H80215	Streptococcus pneumoniae uvs402 protein gene, complete cds	98	174	558
142	3	787	957	gb H80215	Streptococcus pneumoniae uvs402 protein gene, complete cds	100	142	171
142	4	980	3022	gb H80215	Streptococcus pneumoniae uvs402 protein gene, complete cds	95	1997	2043

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
142	5	3020	3595	gb M80215	Streptococcus pneumoniae uvs402 protein gene, complete cds	100	153	576
145	1	1	219	emb Z35135 SPAL	S.pneumoniae alia gene for amIA-like gene A	97	185	219
145	2	171	1994	gb L20556	Streptococcus pneumoniae plpA gene, partial cds	99	1811	1824
145	3	2187	7599	emb Z47210 SPDE	S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs	99	1052	5313
145	4	9934	7766	gb M90527	Streptococcus pneumoniae penicillin-binding protein (ponA) gene, complete cds	99	2169	2169
145	5	10488	9922	gb M90527	Streptococcus pneumoniae penicillin-binding protein (ponA) gene, complete cds	99	512	567
146	1	1159	4	emb Z82002 SP28	S.pneumoniae pcgB and pcgC genes	98	156	156
146	2	344	90	emb Z82002 SP28	S.pneumoniae pcgB and pcgC genes	98	255	255
146	16	11795	10794	emb Z82002 SP28	S.pneumoniae pcgB and pcgC genes	85	276	1002
147	11	10678	10202	emb Z21702 SPUN	S.pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	98	477	477
147	12	11338	10676	emb Z21702 SPUN	S.pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	99	663	663
148	12	9009	8815	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	90	180	195
156	4	1154	1402	emb X63602 SP80	S.pneumoniae msaA-Box	94	185	249
159	13	9048	8521	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	526	528
160	1	1	147	emb Z26851 SPAT	S.pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	100	142	147
160	2	179	898	emb Z26851 SPAT	S.pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	99	720	720
160	3	906	1406	emb Z26850 SPAT	S.pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	95	501	501
160	4	1373	1942	emb Z26850 SPAT	S.pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	87	306	570
161	1	1	984	emb X77249 SPR6	S.pneumoniae (R6) ctaR/ctaH genes	99	984	984
161	7	6910	7497	emb X83917 SPCY	S.pneumoniae orf19yB and gyrB gene encoding DNA gyrase B subunit	99	437	588
161	8	7443	9386	emb X83917 SPCY	S.pneumoniae orf19yB and gyrB gene encoding DNA gyrase B subunit	98	1912	1944
163	1	2	2155	gb L20559	Streptococcus pneumoniae Exp5 gene, partial cds	98	327	2154

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
165	1	32	1618	gb J01796	S.pneumoniae malX and malM genes encoding membrane protein and amylomaltase, complete cds, and malP gene encoding phosphorylase	99	1587	1587
165	2	1608	3902	gb J01796	S.pneumoniae malX and malM genes encoding membrane protein and amylomaltase, complete cds, and malP gene encoding phosphorylase	100	280	2295
166	1	378	4	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	100	375	375
166	2	1507	320	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	1188	1188
166	3	3240	1432	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	563	1809
167	1	1077	328	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	94	155	750
167	2	1844	999	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	98	405	846
167	3	2714	1842	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	97	604	873
167	4	3399	2641	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	99	703	759
168	1	1	2259	gb L20558	Streptococcus pneumoniae Exp4 gene, partial cds	99	282	2259
170	10	7338	7685	emb Z77726 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	95	315	348
172	6	2462	4981	gb U47625	Streptococcus pneumoniae formate acetyltransferase (exp72) gene, partial cds	97	365	2520
175	1	373	20	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	89	353	354
175	4	1843	3621	emb Z47210 SPDE	S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs	95	89	1779
176	5	3984	2980	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	573	1005
178	1	3	425	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	423	423
179	1	426	70	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	99	338	357
180	3	3084	1855	emb X95718 SPGY	S.pneumoniae gyrA gene	99	381	1230
186	1	714	4	emb Z79691 SOOR	S.pneumoniae ynf(A,B,C,D,E), ftsL, pbpX and regR genes	98	59	711
186	2	2254	608	emb Z79691 SOOR	S.pneumoniae ynf(A,B,C,D,E), ftsL, pbpX and regR genes	98	315	1647
186	3	707	880	emb Z79691 SOOR	S.pneumoniae ynf(A,B,C,D,E), ftsL, pbpX and regR genes	98	174	174
189	1	2	259	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	258	258
189	2	600	385	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	98	204	216

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
189	3	1018	851	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	99	168	168
189	4	1012	2154	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	99	1062	1143
191	9	7829	7524	emb X63602 SP80	S.pneumoniae mmsA-Box	95	234	306
194	1	1	729	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	91	728	729
199	2	1117	881	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	96	211	237
199	4	1499	1762	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	89	248	264
199	5	1781	2284	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	98	504	504
203	1	1977	337	gb L20563	Streptococcus pneumoniae Exp9 gene, partial cds	99	342	1641
204	1	1145	3	gb L36131	Streptococcus pneumoniae exp10 gene, complete cds, recA gene, 5' end	99	1143	1143
208	1	59	2296	gb U89711	Streptococcus pneumoniae pneumococcal surface protein A PspA (pspA) gene, complete cds	90	471	2238
213	3	2455	2123	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	96	332	333
216	1	368	12	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	99	338	357
216	3	2650	2327	gb M28678	S.pneumoniae promoter sequence DNA	98	86	324
222	1	417	4	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	414	414
227	3	5266	4238	emb AJ000336 SP	Streptococcus pneumoniae ldh gene	99	1029	1029
239	1	1	804	gb M31296	S.pneumoniae recP gene, complete cds	95	484	804
247	3	1625	1807	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	178	183
249	3	921	1364	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	443	444
253	1	362	3	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	360	360
253	5	1238	2050	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	95	420	813

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
253	6	2069	2572	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	97	504	504
255	1	3	800	emb 282002 SP28	S.pneumoniae pcgB and pcgC genes	97	531	798
255	2	798	1841	emb 282002 SP28	S.pneumoniae pcgB and pcgC genes	97	672	1044
255	3	2493	1969	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	92	435	525
257	2	985	770	emb X17337 SPAM	Streptococcus pneumoniae <i>ami</i> locus conferring aminopterin resistance	96	117	216
257	3	1245	907	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	339	339
267	2	495	1208	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	95	84	714
267	3	1291	2277	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	97	755	987
267	4	2261	3601	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	98	1341	1341
267	5	3561	4136	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	576	576
267	6	4164	4949	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	748	786
267	7	5544	5140	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	100	186	405
268	4	1793	1990	emb X63602 SP90	S.pneumoniae <i>mmsA</i> -Box	89	194	198
271	1	562	104	gb M29686	S.pneumoniae mismatch repair (hexB) gene, complete cds	93	160	459
291	1	75	524	gb U04047	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	450	450
291	2	1001	525	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	87	205	477
291	3	807	559	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	90	170	249
291	4	1374	1099	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	85	264	276

TABLE I *S. pneumoniae* - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
293	1	3	1673	emb Z67740 SPGY	<i>S. pneumoniae</i> gyrB gene and unknown orf	98	553	1671
296	1	1434	151	emb Z47210 SPDE	<i>S. pneumoniae</i> dexB, cap3A, cap3B and cap3C genes and orfs	99	430	1284
317	1	157	510	emb Z67739 SPPA	<i>S. pneumoniae</i> parC, parE and transposase genes and unknown orf	89	353	354
325	2	1237	485	emb Z83335 SP28	<i>S. pneumoniae</i> dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	91	299	753
326	1	1	462	emb Z82001 SP28	<i>S. pneumoniae</i> pcpA gene and open reading frames	100	233	462
327	1	603	64	emb Z83335 SP28	<i>S. pneumoniae</i> dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	89	540
334	1	153	545	gb U41735	<i>Streptococcus pneumoniae</i> peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	87	91	393
336	1	308	93	emb Z26850 SPAT	<i>S. pneumoniae</i> (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	97	102	216
360	1	1	519	emb Z67739 SPPA	<i>S. pneumoniae</i> parC, parE and transposase genes and unknown orf	95	435	519
360	4	1598	1960	emb Z83335 SP28	<i>S. pneumoniae</i> dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	353	363
362	1	673	2	emb Z83335 SP28	<i>S. pneumoniae</i> dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	95	63	672
362	2	1168	728	gb U04047	<i>Streptococcus pneumoniae</i> SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	441	441
384	1	347	111	emb X85787 SPCP	<i>S. pneumoniae</i> dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, cps14M genes	94	54	237

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
228	2	1760	1942	pir F60663 F606	translation elongation factor Tu - Streptococcus oralis	100	100	183
319	1	2	205	gi 984927	neomycin phosphotransferase (cloning vector pBSL99)	100	100	204
260	1	2	1138	pir F60663 F606	translation elongation factor Tu - Streptococcus oralis	99	98	1137
25	2	486	1394	gi 1574495	hypothetical (Haemophilus influenzae)	98	96	909
94	2	685	1002	gi 310627	phosphoenolpyruvate:sugar phosphotransferase system HPr (Streptococcus mutans)	98	93	318
312	1	190	2	gi 347999	ATP-dependent protease proteolytic subunit (Streptococcus salivarius)	98	95	189
329	1	1	807	gi 924848	inosine monophosphate dehydrogenase (Streptococcus pyogenes)	98	94	807
336	2	290	589	gi 987050	lacZ gene product (unidentified cloning vector)	98	98	300
181	9	5948	7366	gi 153755	phospho-beta-D-galactosidase (EC 3.2.1.85) (Lactococcus lactis cremoris)	97	94	1419
312	2	1044	361	gi 347998	uracil phosphoribosyltransferase (Streptococcus salivarius)	97	88	684
32	8	6575	7466	sp P37214 ERA_S	GTP-BINDING PROTEIN ERA HOMOLOG	96	91	912
94	3	951	2741	gi 153615	phosphoenolpyruvate:sugar phosphotransferase system enzyme I (Streptococcus salivarius)	96	92	1791
127	1	1	168	gi 581299	initiation factor IF-1 (Lactococcus lactis)	96	89	168
128	14	10438	11154	gi 1276873	DeoD (Streptococcus thermophilus)	96	93	717
181	4	1362	1598	gi 46606	lacD polypeptide (AA 1-326) (Staphylococcus aureus)	96	80	237
218	1	1	834	gi 1743856	intragenic coaggregation-relevant adhesin (Streptococcus gordonii)	96	93	834
319	2	115	441	gi 208225	heat-shock protein 82/neomycin phosphotransferase fusion protein (hsp82-neo) (unidentified cloning vector)	96	96	327
54	12	8622	10967	gnl PID d100972	pyruvate formate-lyase (Streptococcus mutans)	95	89	2346
181	2	606	1289	gi 149396	lacD (Lactococcus lactis)	95	89	684
46	3	3410	3045	gi 1850606	YlxM (Streptococcus mutans)	94	86	366
89	10	7972	7337	gi 703442	thymidine kinase (Streptococcus gordonii)	94	86	636
148	9	6431	7354	gi 995767	UDP-glucose pyrophosphorylase (Streptococcus pyogenes)	94	85	924
160	7	4430	5848	gi 153573	H+ ATPase (Enterococcus faecalis)	94	87	1419
2	3	4598	3513	gi 153763	plasma receptor (Streptococcus pyogenes)	93	86	1086
12	8	7877	6204	gi 1103865	formyl-tetrahydrofolate synthetase (Streptococcus mutans)	93	84	1674

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
65	11	4734	5120	gi 40150	L14 protein (AA 1-122) [Bacillus subtilis]	93	87	387
68	1	53	1297	gi 47341	antitumor protein [Streptococcus pyogenes]	93	87	1245
80	1	3	299	gnl pid d101166	ribosomal protein S7 [Bacillus subtilis]	93	84	297
127	3	695	1093	gi 142462	ribosomal protein S11 [Bacillus subtilis]	93	86	399
160	5	1924	3462	gi 1773264	ATPase, alpha subunit [Streptococcus mutans]	93	85	1539
211	5	3757	3047	gi 535273	aminopeptidase C [Streptococcus thermophilus]	93	82	711
262	1	16	564	gi 149394	lacB [Lactococcus lactis]	93	90	549
366	1	197	3	gi 295259	tryptophan synthase beta subunit [Synechocystis sp.]	93	91	195
25	3	1392	1976	gi 1574496	hypothetical [Haemophilus influenzae]	92	80	585
36	21	20781	19927	gi 310632	hydrophobic membrane protein [Streptococcus gordonii]	92	86	855
181	3	1265	1534	gi 149396	lacD [Lactococcus lactis]	92	83	270
181	7	3662	4060	gi 149410	enzyme III [Lactococcus lactis]	92	83	399
32	4	5631	3937	gnl pid e294090	fibronectin-binding protein-like protein A [Streptococcus gordonii]	91	85	1695
46	2	3054	1462	gi 1850607	signal recognition particle Ffh [Streptococcus mutans]	91	84	1593
65	10	4442	4726	pir S17865 S178	ribosomal protein S17 - Bacillus stearothermophilus	91	80	285
77	2	260	1900	gi 287871	groEL gene product [Lactococcus lactis]	91	82	1641
84	1	2	2056	gi 871784	Cip-like ATP-dependent protease binding subunit [Bos taurus]	91	79	2055
99	8	10750	9272	gi 153740	sucrose phosphorylase [Streptococcus mutans]	91	84	1479
99	9	11947	11072	gi 153739	membrane protein [Streptococcus mutans]	91	78	876
127	5	2065	2469	pir S07223 R595	ribosomal protein L17 - Bacillus stearothermophilus	91	78	405
132	6	9539	9390	gi 143065	hubst [Bacillus stearothermophilus]	91	89	150
137	8	4765	6153	gnl pid d100347	Na ⁺ -ATPase beta subunit [Enterococcus hirae]	91	79	1389
151	7	11119	9734	gi 1815634	glutamine synthetase type 1 [Streptococcus agalactiae]	91	82	1386
201	2	1798	278	gi 2208998	dextran glucosidase DEXS [Streptococcus suis]	91	79	1521
222	2	673	1839	gi 153741	ATP-binding protein [Streptococcus mutans]	91	85	1167
293	5	4113	4400	gi 1196921	unknown protein [insertion sequence IS861]	91	71	288
32	7	6166	6570	pir A36933 A369	diacylglycerol kinase homolog - Streptococcus mutans	90	77	405

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
33	2	841	527	gi 1196921	unknown protein [Insertion sequence IS861]	90	70	315
48	27	20908	19757	gnl pid e274705	lactate oxidase [Streptococcus iniae]	90	80	1152
55	21	19777	18515	gnl pid e221213	CipX protein [Bacillus subtilis]	90	75	1263
56	2	717	977	gi 1710133	flagellar filament cap [Borrelia burgdorferi]	90	50	261
65	1	1	606	gi 1165303	LJ [Bacillus subtilis]	90	75	606
114	1	2	988	gi 153562	aspartate beta-semialdehyde dehydrogenase (EC 1.2.1.11) [Streptococcus mutans]	90	80	987
120	1	1345	827	gi 407880	ORP1 [Streptococcus equisimilis]	90	75	519
159	12	7690	8298	gi 143012	GMP synthetase [Bacillus subtilis]	90	84	609
166	4	4076	3282	gi 1661179	high affinity branched chain amino acid transport protein [Streptococcus mutans]	90	78	795
183	1	28	1395	gi 308858	ATP:pyruvate 2-O-phosphotransferase [Lactococcus lactis]	90	76	1368
191	3	2891	1662	gi 149521	tryptophan synthase beta subunit [Lactococcus lactis]	90	78	1230
198	2	1551	436	gi 2323342	[AF014460] CcpA [Streptococcus mutans]	90	76	1116
305	1	37	783	gi 1573551	asparagine synthetase A (asnA) [Haemophilus influenzae]	90	80	747
8	3	2285	3343	gi 149434	putative [Lactococcus lactis]	89	78	1059
46	8	7577	7362	pir A45434 A454	ribosomal protein L19 - Bacillus stearothermophilus	89	76	216
49	9	8363	10342	gi 153792	recP peptide [Streptococcus pneumoniae]	89	83	1980
51	14	18410	19447	gi 308857	ATP-D-fructose 6-phosphate 1-phosphotransferase [Lactococcus lactis]	89	81	1038
57	11	9686	10669	gnl pid d100932	H2O-forming NADH Oxidase [Streptococcus mutans]	89	77	984
65	5	2418	2786	gi 1165307	S19 [Bacillus subtilis]	89	81	369
65	8	3806	4225	sp P14577 RL16_50S	RIBOSOMAL PROTEIN L16	89	82	420
65	18	8219	8719	gi 143417	ribosomal protein S5 [Bacillus stearothermophilus]	89	76	501
73	9	6337	5315	gi 532204	prf [Listeria monocytogenes]	89	70	1023
76	3	3360	1465	gnl pid e200671	lepA gene product [Bacillus subtilis]	89	76	1896
99	10	12818	11919	gi 153738	membrane protein [Streptococcus mutans]	89	73	900
120	2	3552	1300	gi 407881	stringent response-like protein [Streptococcus equisimilis]	89	79	2253
122	5	4512	2791	gnl pid e280490	unknown [Streptococcus pneumoniae]	89	81	1722

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
176	1	669	4	gi 47394	5-oxopropyl-peptidase [Streptococcus pyogenes]	89	78	666
177	6	3050	3934	gi 912423	putative [Lactococcus lactis]	89	71	885
181	8	4033	5751	gi 149411	enzyme III [Lactococcus lactis]	89	80	1719
211	4	3149	2793	gi 535273	aminopeptidase C [Streptococcus thermophilus]	89	83	357
361	1	431	838	gi 1196922	unknown protein [Insertion sequence IS861]	89	70	408
34	17	11839	10535	sp P30053 SYH_S	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE-TRNA LIGASE) (HISRS)	88	78	1305
38	3	1646	2623	gi 2058544	putative ABC transporter subunit ComYA [Streptococcus gordonii]	88	78	978
54	1	3	227	gnl PID d101320	Yqgu [Bacillus subtilis]	88	66	225
57	2	611	1468	gnl PID e134943	putative reductase I [Saccharomyces cerevisiae]	88	75	858
65	13	5497	6069	pir A29102 R5B5	ribosomal protein L5 - Bacillus stearothermophilus	88	75	573
65	20	9030	9500	gi 2078381	ribosomal protein L15 [Staphylococcus aureus]	88	83	471
78	3	3636	1108	gnl PID d100781	lysyl-aminopeptidase [Lactococcus lactis]	88	80	2529
106	12	12965	12054	gi 2407215	[AF017421] putative heat shock protein HtpX [Streptococcus gordonii]	88	72	912
107	2	219	962	gnl PID e339862	putative acylneuraminidase lyase [Clostridium tertium]	88	75	744
111	8	14073	10420	gi 402363	RNA polymerase beta-subunit [Bacillus subtilis]	88	74	3654
126	9	13096	12062	gnl PID e311468	unknown [Bacillus subtilis]	88	74	1035
140	17	19143	18874	gi 1573659	H. influenzae predicted coding region HI0659 [Haemophilus influenzae]	88	61	270
144	1	394	555	gnl PID e274705	lactate oxidase [Streptococcus iniae]	88	75	162
148	4	2723	3493	gi 1591672	phosphate transport system ATP-binding protein [Methanococcus jannaschii]	88	68	771
160	8	5853	6278	gi 1773267	ATPase, epsilon subunit [Streptococcus mutans]	88	65	426
177	4	1770	2885	gi 149436	putative [Lactococcus lactis]	88	72	1116
211	6	4140	3613	gi 535273	aminopeptidase C [Streptococcus thermophilus]	88	74	528
231	4	580	957	gi 40186	homologous to E.coli ribosomal protein L27 [Bacillus subtilis]	88	78	378
260	5	2387	2998	gi 1196922	unknown protein [Insertion sequence IS861]	88	69	612
291	6	2017	3375	gnl PID d100571	adenylosuccinate synthetase [Bacillus subtilis]	88	75	1359
319	4	658	317	gi 603578	serine/threonine kinase [Phytophthora capsici]	88	88	342
40	5	4353	4514	gi 153672	lactose repressor [Streptococcus mutans]	87	56	162

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
49	10	10660	10929	gi 1196921	unknown protein [Insertion sequence IS861]	87	72	270
65	7	3140	3808	gi 1165309	S3 [Bacillus subtilis]	87	73	669
65	15	6623	7039	gi 1044978	ribosomal protein S8 [Bacillus subtilis]	87	73	417
75	8	5411	6625	gi 1877422	galactokinase [Streptococcus mutans]	87	78	1215
80	2	703	2805	gnl PID d101166	elongation factor G [Bacillus subtilis]	87	76	2103
82	1	541	248	gi 1196921	unknown protein [Insertion sequence IS861]	87	69	294
140	23	25033	23897	gnl PID e254999	phenylalanyl-tRNA synthetase beta subunit [Bacillus subtilis]	87	74	1137
214	14	10441	8516	gi 2281305	glucose inhibited division protein homolog Gida [Lactococcus lactis cremoris]	87	75	1926
220	2	2742	874	gnl PID e324358	product highly similar to elongation factor EF-G [Bacillus subtilis]	87	73	1869
260	4	2096	2389	gi 1196921	unknown protein [Insertion sequence IS861]	87	72	294
323	1	27	650	gi 897795	30S ribosomal protein (pediococcus acidilactici)	87	73	624
357	1	154	570	gi 1044978	ribosomal protein S8 [Bacillus subtilis]	87	73	417
49	11	10927	11445	gi 1196922	unknown protein [Insertion sequence IS861]	86	63	519
59	12	7461	9224	gi 951051	relaxase [Streptococcus pneumoniae]	86	68	1764
65	4	1553	2401	pir A03759 R5BS	ribosomal protein L2 - Bacillus stearothermophilus	86	77	849
65	23	10957	11610	gi 44074	adenylate kinase [Lactococcus lactis]	86	76	654
82	4	4374	4856	gi 153745	mannitol-specific enzyme III [Streptococcus mutans]	86	72	483
102	4	4270	4986	gnl PID e264705	OMP decarboxylase [Lactococcus lactis]	86	76	717
106	6	7824	6880	gnl PID e137598	aspartate transcarbamylase [Lactobacillus leichmannii]	86	68	945
107	1	1	273	gnl PID e339862	putative acylneuraminate lyase [Clostridium tertium]	86	71	273
111	7	10432	6710	gnl PID e228283	DNA-dependent RNA polymerase [Streptococcus pyogenes]	86	80	3723
131	9	5704	4892	gi 1661193	polipoprotein diacylglycerol transferase [Streptococcus mutans]	86	71	813
134	7	6430	7980	gi 2388637	glycerol kinase [Enterococcus faecalis]	86	73	1551
146	11	7473	6583	gi 1591731	melvalonate kinase [Methanococcus jannaschii]	86	72	891
153	2	595	2010	gi 2160707	dipeptidase [Lactococcus lactis]	86	78	1416
154	1	2	1435	gi 1857246	6-phosphogluconate dehydrogenase [Lactococcus lactis]	86	74	1434

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
161	5	5025	6284	gi 47529	Unknown [Streptococcus salivarius]	85	66	1260
184	1	2	1493	gi 642667	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase [Streptococcus mutans]	86	73	1482
210	8	3659	6571	gi 153661	translational initiation factor IF2 [Enterococcus faecium]	86	76	2913
250	1	2	187	gi 1573551	asparagine synthetase A (asna) [Haemophilus influenzae]	86	68	186
36	4	2644	3909	gi 2149909	cell division protein [Enterococcus faecalis]	85	73	1266
38	4	2475	3587	gi 2058545	putative ABC transporter subunit ComYB [Streptococcus gordonii]	85	72	1113
38	5	3577	3915	gi 2058546	ComYC [Streptococcus gordonii]	85	80	339
57	5	2797	3789	gnl PID d101316	YqfJ [Bacillus subtilis]	85	72	993
82	5	4915	6054	gi 153746	mannitol-phosphate dehydrogenase [Streptococcus mutans]	85	68	1140
83	15	14690	15793	gi 143371	phosphoribosyl aminimidazole synthetase (PUR-M) [Bacillus subtilis]	85	69	1104
87	2	1417	2388	gi 1184967	SerR [Streptococcus mutans]	85	69	972
108	3	2666	3154	gi 153566	ORF (19K protein) [Enterococcus faecalis]	85	67	489
127	2	312	692	gi 1044989	ribosomal protein S13 [Bacillus subtilis]	85	72	381
128	3	1534	2409	gi 1685110	tetrahydrofolate dehydrogenase/cyclohydrolase [Streptococcus thermophilus]	85	71	876
137	7	2982	4767	gnl PID d100347	Na ⁺ -ATPase alpha subunit [Enterococcus hirae]	85	74	1806
170	2	2622	703	gnl PID d102006	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN E. COLI, H. INFLUENZAE AND NEISSERIA MENINGITIDIS. [Bacillus subtilis]	85	70	1914
187	5	3760	4386	gi 727436	putative 20-kDa protein [Lactococcus lactis]	85	65	627
233	2	728	1873	gi 1163116	ORF-5 [Streptococcus pneumoniae]	85	67	1146
234	3	962	1255	gi 2293155	(AF008220) Ytia [Bacillus subtilis]	85	61	294
240	1	309	1931	gi 143597	CTP synthetase [Bacillus subtilis]	85	70	1623
6	1	199	1521	gi 508979	GTP-binding protein [Bacillus subtilis]	84	72	1323
10	4	4375	3443	gnl PID e139862	putative acylneuraminase lyase [Clostridium tertium]	84	70	933
14	1	63	2093	gi 520753	DNA topoisomerase I [Bacillus subtilis]	84	69	2031
19	4	1793	2593	gi 2352484	(AF005098) RNASEH II [Lactococcus lactis]	84	68	801
20	17	17720	19687	gnl PID d100584	cell division protein [Bacillus subtilis]	84	71	1968
22	28	21723	20884	gi 299163	alanine dehydrogenase [Bacillus subtilis]	84	68	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
30	10	7730	6792	gi PID d100296	fructokinase [Streptococcus mutans]	84	75	939
33	9	5650	5300	gi 147194	pna protein [Escherichia coli]	84	71	351
36	22	21551	20772	gi 310631	ATP binding protein [Streptococcus gordonii]	84	72	780
48	4	2837	2505	gi 882609	6-phospho-beta-glucosidase [Escherichia coli]	84	69	333
58	1	41	1516	gi 450849	amylase [Streptococcus bovis]	84	73	1476
59	10	6715	7116	gi 951053	ORF10, putative [Streptococcus pneumoniae]	84	74	402
62	1	21	644	gi 806487	ORF211; putative [Lactococcus lactis]	84	66	624
65	17	7779	8207	gi 1044980	ribosomal protein L18 [Bacillus subtilis]	84	73	429
65	21	9507	10397	gi 44073	SecY protein [Lactococcus lactis]	84	68	891
106	4	5474	2262	gi PID e19387	carbamoyl-phosphate synthase [Lactobacillus plantarum]	84	73	3213
139	1	147	4	gi 806487	ORF211; putative [Lactococcus lactis]	84	63	144
163	4	4690	5910	gi 2293164	(AF008220) SAM synthase [Bacillus subtilis]	84	69	1221
192	1	46	1308	gi 495046	tripeptidase [Lactococcus lactis]	84	73	1263
348	1	671	6	gi 1787753	(AF000245) f346; 79 pct identical to 336 amino acids of ADH1_2YHNO SW: P20368 but has 10 additional N-ter residues [Escherichia coli]	84	71	666
3	4	1572	3575	gi 143766	(thrSv) (EC 6.1.1.3) [Bacillus subtilis]	83	65	2004
9	6	3893	3417	gnl PID d100576	single strand DNA binding protein [Bacillus subtilis]	83	68	477
17	15	7426	8457	gi 520738	comA protein [Streptococcus pneumoniae]	83	66	1032
20	12	13860	14144	gnl PID d100583	unknown [Bacillus subtilis]	83	61	285
23	4	3358	2606	gi 1788294	(AF000290) o236; This 238 aa orf is 40 pct identical (5 gaps) to 231 residues of an approx. 248 aa protein YEB_C001 SW: P24237 [Escherichia coli]	83	74	753
28	6	3304	3005	gi 1573659	H. influenzae predicted coding region HI0659 [Haemophilus influenzae]	83	57	300
35	7	5108	3867	gi 111707	hypothetical nucleotide binding protein [Acholeplasma laidlawii]	83	63	1242
55	19	17932	17528	gi 537085	ORF f141 [Escherichia coli]	83	59	405
55	20	18539	17919	gi 496558	orfX [Bacillus subtilis]	83	69	621
65	6	2795	3142	gi 1165308	L22 [Bacillus subtilis]	83	64	348
68	6	6877	6683	gi 1213494	immunoglobulin A1 protease [Streptococcus pneumoniae]	83	54	195

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
87	15	15112	14771	[gnl pid e323522]	putative rpoZ protein [Bacillus subtilis]	83	54	342
96	12	8963	9631	[gi 47394]	[5-oxopropyl-L-peptidase [Streptococcus pyogenes]	83	73	669
98	1	3	263	[gi 1183885]	[glutamine-binding subunit [Bacillus subtilis]	83	55	261
120	4	7170	5233	[gi 310630]	[zinc metalloprotease [Streptococcus gordonii]	83	72	1938
127	7	2998	4347	[gi 1500567]	[M. jannaschii predicted coding region MJ1665 [Methanococcus jannaschii]	83	72	1350
137	1	3	440	[gi 472918]	[v-type Na-ATPase [Enterococcus hirae]	83	60	438
160	6	3466	4356	[gi 1773265]	[ATPase, gamma subunit [Streptococcus mutans]	83	67	891
214	4	2278	2964	[gi 663279]	[transposase [Streptococcus pneumoniae]	83	72	687
226	3	2367	2020	[gi 142154]	[thioredoxin [Synechococcus PCC6301]	83	58	348
303	1	3	1049	[gi 40046]	[phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus]	83	67	777
303	2	1155	1931	[gi 289282]	[glutaryl-tRNA synthetase [Bacillus subtilis]	82	64	1053
6	17	15370	14318	[gi 633147]	[ribose-phosphate pyrophosphokinase [Bacillus caldolyticus]	82	69	204
7	1	299	96	[gi 143668]	[ribosomal protein L28 [Bacillus subtilis]	82	46	390
9	3	1479	1090	[gi 385178]	[unknown [Bacillus subtilis]	82	60	315
9	7	4213	3899	[gnl pid d100576]	[ribosomal protein S6 [Bacillus subtilis]	82	68	747
12	6	4688	3942	[gnl pid d100571]	[unknown [Bacillus subtilis]	82	69	1416
22	17	13422	14837	[gi 520754]	[putative [Bacillus subtilis]	82	62	762
22	18	14897	15658	[gnl pid d101929]	[uridine monophosphate kinase [Synechocystis sp.]	82	68	831
33	16	11471	10641	[gnl pid d101190]	[ORF4 [Streptococcus mutans]	82	68	831
35	9	7400	6255	[gi 1881543]	[UDP-N-acetylglucosamine-2-epimerase [Streptococcus pneumoniae]	82	68	1146
40	10	8003	7533	[gi 1173519]	[riboflavin synthase beta subunit [Actinobacillus pleuropneumoniae]	82	68	471
48	32	23159	23437	[gi 1930092]	[outer membrane protein [Campylobacter jejuni]	82	61	279
52	14	13833	14765	[gi 142521]	[deoxyribodipyrimidine photolyase [Bacillus subtilis]	82	61	933
60	4	4737	1849	[gnl pid d102221]	[AB001610] uvrA [Deinococcus radiodurans]	82	66	2889
62	4	2131	1457	[gi 12246749]	[AF009622] thioredoxin reductase [Listeria monocytogenes]	82	63	675
71	11	16586	17518	[gnl pid g322063]	[ss-1, 4-galactosyltransferase [Streptococcus pneumoniae]	82	60	933
73	13	9222	7837	[gnl pid d100586]	[unknown [Bacillus subtilis]	82	65	1386

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
74	1	1	3771	gnl PID d101199	alkaline anyopolulanase (Bacillus sp.)	82	68	3771
83	9	3696	3983	gnl PID e305362	unnamed protein product (Streptococcus thermophilus)	82	52	288
86	11	10776	9394	gsl 683583	5-enolpyruvylshikimate-3-phosphate synthase (Lactococcus lactis)	82	67	1383
89	12	8295	9752	gsl 40025	homologous to E.coli 50K (Bacillus subtilis)	82	66	1458
115	9	10347	8812	gnl PID d102090	(AB003927) phospho-beta-galactosidase 1 (Lactobacillus gasserii)	82	74	1536
118	1	1	1332	gnl PID d100579	seryl-tRNA synthetase (Bacillus subtilis)	82	71	1332
151	3	4657	6246	pir S06097 S060	type I site-specific deoxyribonuclease (EC 3.1.21.3) CfrA chain S - Citrobacter freundii	82	66	1590
173	6	4183	3503	gsl 2313836	(AE000584) conserved hypothetical protein (Helicobacter pylori)	82	68	681
177	12	5481	7442	gnl PID d101999	(AB001341) NcrB (Escherichia coli)	82	58	1962
193	2	178	576	pir S08564 R3BS	ribosomal protein S9 - Bacillus stearothermophilus	82	70	399
245	2	258	845	gsl 146402	EcoA type I restriction-modification enzyme S subunit (Escherichia coli)	82	68	588
9	5	3400	3146	gnl PID d100576	ribosomal protein S18 (Bacillus subtilis)	81	66	255
16	7	7484	8413	gsl 1100074	tryptophanyl-tRNA synthetase (Clostridium longisporum)	81	70	930
20	11	10308	13820	gnl PID d100583	transcription-repair coupling factor (Bacillus subtilis)	81	63	3513
38	2	1232	1606	gsl 2058543	putative DNA binding protein (Streptococcus gordonii)	81	63	375
45	2	3061	1751	gsl 460259	enolase (Bacillus subtilis)	81	67	1311
46	1	2	1267	gsl 431231	uracil permease (Bacillus caldolyticus)	81	61	1266
48	3	2453	1440	gnl PID d100453	Mannosephosphate isomerase (Streptococcus mutans)	81	70	1014
54	2	1106	336	gsl 154752	transport protein (Agrobacterium tumefaciens)	81	64	771
65	22	10306	10821	gsl 44073	SecY protein (Lactococcus lactis)	81	66	516
89	4	3874	2603	gsl 556886	serine hydroxymethyltransferase (Bacillus subtilis)	81	69	1272
99	16	19126	18929	gsl 2313526	(AE000557) H. pylori predicted coding region HP0611 (Helicobacter pylori)	81	75	198
106	7	8373	7822	gnl PID e199384	pyrR (Lactobacillus plantarum)	81	61	552
108	6	5054	6877	gsl 1469939	group B oligopeptidase PepB (Streptococcus agalactiae)	81	66	1824
113	15	15899	18283	pir S09411 S094	spoIIIE protein - Bacillus subtilis	81	65	2385
128	5	3359	3634	gsl 1685111	orf1091 (Streptococcus thermophilus)	81	69	276

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
151	1	830	3211	gi 304896	EcoE type I restriction-modification enzyme R subunit (Escherichia coli)	81	59	2382
159	11	6722	7837	gi 2233288	GMP synthetase (Bacillus subtilis)	81	69	1116
170	1	739	458	gnl PID d102006	(AB001488) FUNCTION UNKNOWN. (Bacillus subtilis)	81	55	282
191	2	1759	893	gi 149522	tryptophan synthase alpha subunit (Lactococcus lactis)	81	65	867
214	3	2290	1994	gi 157587	reverse transcriptase endonuclease (Drosophila virilis)	81	43	297
217	4	4415	4008	gi 466473	cellulose phosphorylase enzyme II' (Bacillus stearothermophilus)	81	59	408
262	2	569	868	gi 153675	tagatase 6-P kinase (Streptococcus mutans)	81	68	300
299	1	663	4	gnl PID e301154	StySKI methylase (Salmonella enterica)	81	60	660
366	2	376	83	gi 149521	tryptophan synthase beta subunit (Lactococcus lactis)	81	65	294
12	10	8766	9242	gi 1216490	DNA/pantothenate metabolism flavoprotein (Streptococcus mutans)	80	64	477
17	11	6050	5748	gnl PID e305362	unnamed protein product (Streptococcus thermophilus)	80	67	303
17	16	8455	9066	gi 703126	Leucocin A translocator (Leuconostoc gelidum)	80	59	612
18	3	2440	1613	gi 1591672	phosphate transport system ATP-binding protein (Methanococcus jannaschii)	80	58	828
27	3	4248	1579	gi 452309	valyl-CRNA synthetase (Bacillus subtilis)	80	69	2670
28	7	3671	3288	gi 1573660	H. influenzae predicted coding region H10660 (Haemophilus influenzae)	80	63	384
32	2	902	1933	gnl PID e264499	dihydroorotate dehydrogenase B (Lactococcus lactis)	80	66	1032
39	1	1	1266	gnl PID e234078	hom (Lactococcus lactis)	80	63	1266
52	5	4363	3593	gi 1183884	ATP-binding subunit (Bacillus subtilis)	80	57	771
54	5	4550	4744	gi 2198820	(AF004225) Cux/CDP homeoprotein (Mus musculus)	80	60	195
59	11	7109	7486	gi 951052	ORF9, putative (Streptococcus pneumoniae)	80	68	378
65	3	1230	1550	pir A02815 R585	ribosomal protein L23 - Bacillus stearothermophilus	80	69	321
65	12	5174	5503	pir A02819 R585	ribosomal protein L24 - Bacillus stearothermophilus	80	70	330
66	9	9884	10687	gi 2313836	(AE000584) conserved hypothetical protein (Helicobacter pylori)	80	66	804
82	2	648	2438	gi 622991	mannitol transport protein (Bacillus stearothermophilus)	80	65	1791
85	1	950	630	gi 528995	polyketide synthase (Bacillus subtilis)	80	46	321
89	8	6870	5779	gi 853776	peptide chain release factor 1 (Bacillus subtilis)	80	63	1092
93	12	8718	7438	gnl PID d101959	hypothetical protein (Synechocystis sp.)	80	60	1281

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	5	6854	5751	gnl PID e199386	glutaminase of carbamoyl-phosphate synthase [Lactobacillus plantarum]	80	65	1104
109	2	2160	1450	gi 40056	phoP gene product [Bacillus subtilis]	80	59	711
124	9	4246	3953	gnl PID d102254	30S ribosomal protein S16 [Bacillus subtilis]	80	65	294
128	8	5148	6428	gi 2281308	phosphoenolpyruvate carboxylase [Lactococcus lactis cremoris]	80	66	1281
137	19	12665	11376	gi 159109	NADP-dependent glutamate dehydrogenase [Giardia intestinalis]	80	68	1290
140	19	19699	19457	gi 517210	putative transposase [Streptococcus pyogenes]	80	70	243
158	2	2474	984	gi 1877423	galactose-1-P-uridylyl transferase [Streptococcus mutans]	80	65	1491
171	10	7474	7728	gi 397800	cyclophilin C-associated protein [Mus musculus]	80	60	255
181	1	2	619	gi 149395	lacC [Lactococcus lactis]	80	66	618
333	1	27	539	gi 143467	ribosomal protein S4 [Bacillus subtilis]	80	70	513
329	2	1652	858	gi 533080	RacF protein [Streptococcus pyogenes]	80	63	795
371	1	2	958	gi 442360	ClpC adenosine triphosphatase [Bacillus subtilis]	80	58	957
8	7	4312	5580	gi 149435	putative [Lactococcus lactis]	79	64	1269
23	1	1175	135	gi 1542975	AbcB [Thermoanaerobacterium thermosulfurigenes]	79	61	1041
33	14	9244	8201	gnl PID e253891	UDP-glucose 4-epimerase [Bacillus subtilis]	79	62	1044
36	3	1242	2633	gnl PID e324218	ftsA [Enterococcus hirae]	79	58	1392
38	13	7155	8378	gi 405134	acetate kinase [Bacillus subtilis]	79	58	1224
55	7	9011	8229	gi 1146234	dihydrodipicolinate reductase [Bacillus subtilis]	79	56	783
65	19	8661	8915	gi 2078380	ribosomal protein L30 [Staphylococcus aureus]	79	68	255
69	4	3678	2128	gnl PID e311452	unknown [Bacillus subtilis]	79	64	1551
69	9	7881	7279	gi 677850	hypothetical protein [Staphylococcus aureus]	79	59	603
72	10	8491	9783	gnl PID d101091	hypothetical protein [Synechocystis sp.]	79	62	1293
80	3	2906	7300	gi 143342	polymerase III [Bacillus subtilis]	79	65	4395
82	14	13326	15689	gnl PID e25093	hypothetical protein [Bacillus subtilis]	79	65	2364
86	13	12233	11118	gi 683582	prophenate dehydrogenase [Lactococcus lactis]	79	58	1116
92	3	940	1734	gi 537286	triosephosphate isomerase [Lactococcus lactis]	79	65	795
98	6	4023	4742	gnl PID d100262	LivG protein [Salmonella typhimurium]	79	63	720

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
99	12	16115	14150	gi153736	a-galactosidase (Streptococcus mutans)	79	64	2166
107	7	5684	6406	gi1460080	D-alanine:D-alanine ligase-related protein (Enterococcus faecalis)	79	58	723
113	9	6858	8303	gi1466882	ppg1: B1496_C2_189 (Mycobacterium leprae)	79	64	1446
151	10	13424	12213	gi1450686	3-phosphoglycerate kinase (Thermotoga maritima)	79	60	1212
162	2	1158	1017	gi1506700	CapD (Staphylococcus aureus)	79	67	1860
177	5	2876	3052	gi1912423	putative (Lactococcus lactis)	79	61	177
177	8	4198	4563	gi149429	putative (Lactococcus lactis)	79	61	366
187	3	2728	2907	gnlPID d102002	putative (Lactococcus lactis)	79	53	180
189	7	3589	4350	gnlPID ei183449	putative ATP-binding protein of ABC-type (Bacillus subtilis)	79	61	762
191	5	4249	3449	gi149519	indoleglycerol phosphate synthase (Lactococcus lactis)	79	66	801
211	3	1805	2737	gi1147404	mannose permease subunit II-N-Man (Escherichia coli)	79	57	933
212	3	3863	3621	gnlPID e209004	glutaredoxin-like protein (Lactococcus lactis)	79	58	243
215	1	987	715	gi12293242	(AF008220) arginine succinate synthase (Bacillus subtilis)	79	64	273
323	2	530	781	gi1897795	30S ribosomal protein (Pedococcus acidilactici)	79	67	252
380	1	694	2	gi1184680	polynucleotide phosphorylase (Bacillus subtilis)	79	64	693
384	2	655	239	gi1143328	phoP protein (put.): putative (Bacillus subtilis)	79	59	417
6	3	2820	4091	gi1833767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)	78	62	1272
8	1	50	1786	gi1149432	putative (Lactococcus lactis)	78	63	1737
9	1	351	124	gi1897793	y98 gene product (Pedococcus acidilactici)	78	59	228
15	8	7364	8314	gnlPID d100585	cytateine synthetase A (Bacillus subtilis)	78	63	951
20	10	9738	10310	gnlPID d100583	stage V sporulation (Bacillus subtilis)	78	58	573
20	16	17165	17713	gi149105	hypoxanthine phosphoribosyltransferase (Lactococcus lactis)	78	59	509
22	22	17388	18416	gnlPID d101315	YqfE (Bacillus subtilis)	78	60	1029
22	27	20971	20612	gi1299163	alanine dehydrogenase (Bacillus subtilis)	78	59	360
34	8	7407	7105	gi141015	aspartate-tRNA ligase (Escherichia coli)	78	55	303
35	8	6257	5196	gi11657644	Cap8E (Staphylococcus aureus)	78	60	1062

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
40	11	9287	8001	gi 1173518	GTP cyclohydrolase II/ 3,4-dihydroxy-2-butanone-4-phosphate synthase (Actinobacillus pleuropneumoniae)	78	58	1287
48	31	22422	23183	gi 2314330	(AE000623) glutamine ABC transporter, ATP-binding protein (ginQ) (Helicobacter pylori)	78	58	762
52	2	2101	1430	gi 1183887	Integral membrane protein (Bacillus subtilis)	78	54	672
55	14	13605	12712	gn PID d102026	(AB002150) YbbP (Bacillus subtilis)	78	58	894
55	17	16637	15612	gn PID e313027	hypothetical protein (Bacillus subtilis)	78	51	1026
71	14	19756	19598	gi 179764	calcium channel alpha-1D subunit (Homo sapiens)	78	57	159
74	11	15031	14018	gi 1573279	Holliday junction DNA helicase (ruvB) (Haemophilus influenzae)	78	57	1014
75	9	6623	7972	gi 1877423	galactose-1-P-uridylyl transferase (Streptococcus mutans)	78	62	1350
81	12	12125	13906	gi 1573607	L-fucose isomerase (fucI) (Haemophilus influenzae)	78	66	1782
82	3	2423	4417	gi 153744	ORF X; putative (Streptococcus mutans)	78	64	1995
83	18	16926	18500	gi 143373	phosphoribosyl aminimidazole carboxy formyl formyl transferase/inosine monophosphate cyclohydrolase (PUR-H(J)) (Bacillus subtilis)	78	63	1575
83	20	20212	20775	gi 143364	phosphoribosyl aminimidazole carboxylase I (PUR-E) (Bacillus subtilis)	78	64	564
92	2	165	878	gn PID d101190	ORF2 (Streptococcus mutans)	78	62	714
98	8	5863	6909	gi 2331287	(AF013188) release factor 2 (Bacillus subtilis)	78	63	1047
113	3	1071	2741	gi 580914	dnaX (Bacillus subtilis)	78	64	1671
127	4	1133	2071	gi 142463	RNA polymerase alpha-core-subunit (Bacillus subtilis)	78	59	939
132	1	2782	497	gi 1561763	pullulanase (Bacteroides thetaotaomicron)	78	58	2286
135	4	2698	3537	gi 1788036	(AE000269) NHJ-dependent NAD synthetase (Escherichia coli)	78	66	840
140	24	26853	25423	gi 1100077	phospho-beta-glucosidase (Clostridium longisporum)	78	64	1431
150	5	4690	4514	gi 149464	amino peptidase (Lactococcus lactis)	78	42	177
152	1	1	795	gi 1639915	NADH dehydrogenase subunit (Rhumbardia alata)	78	43	795
162	4	4997	4110	gn PID e323528	putative YhaP protein (Bacillus subtilis)	78	64	888
181	10	8651	7947	gi 149402	lactose repressor (lacR; alt.) (Lactococcus lactis)	78	48	705
200	4	3627	4958	gn PID d100172	Invertase (Zymomonas mobilis)	78	61	1332
203	3	3230	3015	gi 1174237	CycK (Pseudomonas fluorescens)	78	57	216

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
210	9	6789	7172	gi 580902	ORF6 gene product [Bacillus subtilis]	78	42	384
214	6	3810	2797	gnl PID d102049	P. haemolytica O-sialoglycoprotein endopeptidase; P36175 (660) transmembrane [Bacillus subtilis]	78	60	1014
214	13	6322	8163	gi 1377831	unknown [Bacillus subtilis]	78	62	1842
217	1	9	2717	gi 488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	78	64	2709
222	3	2316	3098	gi 1573047	spore germination and vegetative growth protein (gerC2) [Haemophilus influenzae]	78	65	783
268	1	742	8	gi 517210	putative transposase [Streptococcus pyogenes]	78	65	735
276	1	223	753	gnl PID d100306	ribosomal protein L1 [Bacillus subtilis]	78	65	531
312	3	1567	1079	gi 289261	comE ORF2 [Bacillus subtilis]	78	54	489
339	1	117	794	gi 1916729	CadD [Staphylococcus aureus]	78	53	678
342	2	762	265	gi 1842439	phosphatidylglycerophosphate synthase [Bacillus subtilis]	78	59	498
383	1	737	3	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	78	64	735
7	15	11923	11018	gi 1399855	carboxyltransferase beta subunit [Synecococcus PCC7942]	77	63	906
8	2	1698	2255	gi 149433	putative [Lactococcus lactis]	77	59	558
17	14	6948	7550	gi 520738	comA protein [Streptococcus pneumoniae]	77	60	603
30	12	9761	8967	gi 1000451	TreP [Bacillus subtilis]	77	43	795
36	14	11421	12131	gi 1573766	phosphoglyceromutase (gpmA) [Haemophilus influenzae]	77	64	711
55	3	3836	4096	gi 1708640	YeeB [Bacillus subtilis]	77	55	261
61	8	8377	8054	gi 1890649	multidrug resistance protein LmrA [Lactococcus lactis]	77	51	324
65	2	607	1254	gi 40103	ribosomal protein L4 [Bacillus stearothermophilus]	77	63	648
68	8	7509	7240	gi 47551	MRP [Streptococcus suis]	77	68	270
69	1	1083	118	gnl PID e311493	unknown [Bacillus subtilis]	77	57	966
77	5	4583	4026	gnl PID e281578	hypothetical 12.2 kd protein [Bacillus subtilis]	77	60	558
83	14	13104	14552	gi 1590947	amidophosphoribosyltransferase [Methanococcus jannaschii]	77	56	1449
94	4	3006	5444	gnl PID e329895	(A2000496) cyclic nucleotide-gated channel beta subunit [Rattus norvegicus]	77	66	2439
96	11	8518	8880	gi 551879	ORF 1 [Lactococcus lactis]	77	62	363
99	11	14082	12799	gi 153737	sugar-binding protein [Streptococcus mutans]	77	61	1284

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	2	361	1176	gi 148921	LucD protein (Haemophilus influenzae)	77	51	816
108	4	3152	4030	gi 1574730	tellurite resistance protein (tehB) (Haemophilus influenzae)	77	58	879
118	4	3520	3131	gi 1573500	D-alanine permease (daga) (Haemophilus influenzae)	77	57	390
124	4	1796	1071	gi 1573162	tRNA (guanine-N1)-methyltransferase (trmD) (Haemophilus influenzae)	77	58	726
126	4	5909	4614	gnl PID d101163	Srb (Bacillus subtilis)	77	62	1296
128	2	630	1373	gnl PID d101328	Yqiz (Bacillus subtilis)	77	58	744
130	1	1	1287	gnl PID e325013	hypothetical protein (Bacillus subtilis)	77	61	1287
139	5	4388	3639	gi 2291302	(AF008220) YtqA (Bacillus subtilis)	77	59	750
140	11	10931	9582	gi 289284	cysteine-lyase (Bacillus subtilis)	77	64	1350
140	18	19451	19263	gi 517210	putative transposase (Streptococcus pyogenes)	77	66	189
141	2	976	1683	gnl PID e157887	URF5 (aa 1-573) (Drosophila yakuba)	77	50	708
141	4	2735	5293	gi 556258	secA (Listeria monocytogenes)	77	59	2559
144	2	671	2173	gnl PID d100585	lysozyme (Bacillus subtilis)	77	61	1503
163	5	6412	7398	gi 511015	dihydroorotate dehydrogenase A (Lactococcus lactis)	77	62	987
164	10	7841	7074	gnl PID d100964	homologue of iron dicitrate transport ATP-binding protein FecE of E. coli (Bacillus subtilis)	77	52	768
191	8	7257	5791	gi 149516	anthranilate synthase alpha subunit (Lactococcus lactis)	77	57	1467
198	8	5377	5177	gi 1573856	hypothetical (Haemophilus influenzae)	77	66	201
213	1	202	462	gi 1743860	BrcA2 (Mus musculus)	77	50	261
250	2	231	509	gnl PID e334776	YibH protein (Bacillus subtilis)	77	60	279
289	3	1737	1276	gnl PID d100947	Ribosomal protein L10 (Bacillus subtilis)	77	62	462
292	2	1399	668	gi 143004	transfer RNA-Gln synthetase (Bacillus stearothermophilus)	77	58	732
7	3	2734	1166	gnl PID d101824	peptide-chain-release factor 3 (Synechocystis sp.)	76	53	1569
7	23	18474	18235	gi 455157	acyl carrier protein (Cryptomonas phi)	76	57	240
9	8	5706	4342	gi 1146247	asparaginyl-tRNA synthetase (Bacillus subtilis)	76	61	1365
10	5	4531	4385	gnl PID e314495	hypothetical protein (Clostridium perfringens)	76	53	147
18	2	1615	842	gi 1591672	phosphate transport system ATP-binding protein (Methanococcus jannaschii)	76	56	774

TABLE 2
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	37	27796	28173	gnl PID e13389	translation initiation factor IF3 (AA 1-172) [Bacillus stearothermophilus]	76	64	378
35	6	3869	2682	gi 1773346	Cap5G [Staphylococcus aureus]	76	61	1188
48	28	21113	21787	gi 2314328	(AE000623) glutamine ABC transporter, permease protein (glnP) [Helicobacter pylori]	76	52	675
52	12	12881	13786	gi 142521	deoxyribodipyrimidine photolyase [Bacillus subtilis]	76	58	906
55	10	11521	10571	gnl PID e283110	femD [Staphylococcus aureus]	76	61	951
57	8	7824	6559	gi 290561	olr8 [Escherichia coli]	76	47	1266
62	5	2406	2095	gnl PID e313024	hypothetical protein [Bacillus subtilis]	76	59	312
65	9	4223	4441	gi 40148	L29 protein (AA 1-66) [Bacillus subtilis]	76	58	219
68	2	1328	2371	gnl PID e284233	anabolic ornithine carbamoyltransferase [Lactobacillus plantarum]	76	61	1044
69	8	7297	6005	gnl PID d101420	Pyrimidine nucleoside phosphorylase [Bacillus stearothermophilus]	76	61	1293
73	12	7839	7267	gnl PID e243629	unknown [Mycobacterium tuberculosis]	76	53	573
74	5	8433	7039	gnl PID d102048	C. thermocellum beta-glucosidase; P26208 (985) [Bacillus subtilis]	76	60	1395
80	5	7643	7936	gi 2314030	(AE000599) conserved hypothetical protein [Helicobacter pylori]	76	61	294
82	15	16019	16996	gi 1573900	D-alanine permease (degA) [Haemophilus influenzae]	76	56	978
83	19	18616	19884	gi 143374	phosphoribosyl glycylamide synthetase (PUR-D; gtg start codon) [Bacillus subtilis]	76	60	1269
86	14	13409	12231	gi 143806	AroF [Bacillus subtilis]	76	58	1179
87	1	3	1442	gi 153804	sucrose-6-phosphate hydrolase [Streptococcus mutans]	76	59	1440
87	16	15754	15110	gnl PID e323500	putative Gmk protein [Bacillus subtilis]	76	56	645
93	4	1769	1539	gi 1574820	1,4-alpha-glucan branching enzyme (glgB) [Haemophilus influenzae]	76	46	231
94	1	51	365	gi 144313	6.0 kd ORF [Plasmid ColE1]	76	73	315
116	2	2151	1678	gi 153841	pneumococcal surface protein A [Streptococcus pneumoniae]	76	59	474
123	6	3442	5895	gi 1314297	ClpC ATPase [Listeria monocytogenes]	76	59	2454
126	2	2156	2932	gnl PID d101328	Yqj2 [Bacillus subtilis]	76	61	777
128	10	6973	7797	gi 944944	purine nucleoside phosphorylase [Bacillus subtilis]	76	60	825
131	11	6186	5812	gi 1674310	(AE000058) Mycoplasma pneumoniae, MG085 homolog, from M. genitalium [Mycoplasma pneumoniae]	76	47	375

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
139	4	3641	3192	gi 2293302	[AF008220] YtqA [Bacillus subtilis]	76	53	450
140	14	14872	12536	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	76	62	2337
143	2	2593	3905	gi 143795	[transfer RNA-Tyr synthetase [Bacillus subtilis]	76	61	1323
170	6	5095	6114	gnl PID d100959	ycgQ [Bacillus subtilis]	76	44	1020
180	2	1927	557	gi 40019	[ORF 821 (aa 1-821) [Bacillus subtilis]	76	53	1371
191	7	5815	5228	gi 551880	anthranilate synthase beta subunit [Lactococcus lactis]	76	61	588
195	3	3829	2444	gi 2149905	D-glutamic acid adding enzyme [Enterococcus faecalis]	76	60	1386
200	3	1914	3629	gi 431272	lysis protein [Bacillus subtilis]	76	58	1716
201	1	431	207	gi 2208998	dextran glucosidase DEXS [Streptococcus suis]	76	57	225
214	2	1283	2380	gi 663278	transposase [Streptococcus pneumoniae]	76	55	1098
225	3	2338	3411	gi 1552775	ATP-binding protein [Escherichia coli]	76	56	1074
233	1	2	724	gi 1163115	neuraminidase B [Streptococcus pneumoniae]	76	60	723
347	1	523	38	gi 537033	[ORF_f356 [Escherichia coli]	76	60	486
356	2	842	165	gi 2149905	D-glutamic acid adding enzyme [Enterococcus faecalis]	76	61	678
366	3	734	348	gi 149520	[phosphoribosyl anthranilate isomerase [Lactococcus lactis]	76	69	387
5	8	12599	11484	gi 1574293	fimbrial transcription regulation repressor (pilB) [Haemophilus influenzae]	75	61	1116
6	13	12553	11894	gnl PID d102050	lydH [Bacillus subtilis]	75	51	660
9	10	7282	6062	gi 142538	aspartate aminotransferase [Bacillus sp.]	75	55	1221
10	12	8080	7940	gi 149493	SCRPI methylase [Lactococcus lactis]	75	56	141
18	5	4266	3301	gnl PID d101319	YqgH [Bacillus subtilis]	75	52	966
22	4	1838	2728	gi 1373157	orf-X; hypothetical protein; Method: conceptual translation supplied by author [Bacillus subtilis]	75	62	891
30	11	9015	7828	gi 153801	enzyme scr-11 [Streptococcus mutans]	75	64	1188
31	5	2362	2030	gi 2293211	[AF008220] putative thioredoxin [Bacillus subtilis]	75	53	333
32	9	7484	8359	gnl PID d100560	formamidopyrimidine-DNA glycosylase [Streptococcus mutans]	75	61	876
33	4	1735	1448	gi 413976	lipa-52r gene product [Bacillus subtilis]	75	53	288
33	10	6470	5769	gi 533105	unknown [Bacillus subtilis]	75	56	702

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
33	12	6878	7183	pir A00205 FECL	ferredoxin (4Fe-4S) - Clostridium thermaceticum	75	56	306
36	1	181	2	gi 2088739	(AF003141) strong similarity to the FABP/P2/CRBP/CRABP family of transporters [Caenorhabditis elegans]	75	43	180
38	22	14510	15379	gi 1574058	hypothetical [Haemophilus influenzae]	75	56	870
48	33	23398	24066	gi 1930092	outer membrane protein [Campylobacter jejuni]	75	56	669
51	1	2	319	gi 43985	nifS-like gene [Lactobacillus delbrueckii]	75	55	318
51	10	8318	11683	gi 537192	CG Site No. 620; alternate gene names ha, hsp, her, rmx apparent frameshift in GenBank Accession Number X06545 [Escherichia coli]	75	50	3366
54	18	19566	20759	gi 666069	orf2 gene product [Lactobacillus leichmannii]	75	58	1194
57	9	8448	7822	gi 290561	tol88 [Escherichia coli]	75	50	627
65	14	6072	6356	gi 606241	30S ribosomal subunit protein S14 [Escherichia coli]	75	64	285
70	4	3071	2472	gi 1256617	adenine phosphoribosyltransferase [Bacillus subtilis]	75	57	600
71	24	30399	29404	gi 1574390	C4-dicarboxylate transport protein [Haemophilus influenzae]	75	57	996
73	2	910	455	gnl PID e249656	Yner [Bacillus subtilis]	75	57	456
79	1	1810	491	gi 1146219	28.2% of identity to the Escherichia coli GTP-binding protein Era; putative [Bacillus subtilis]	75	59	1320
82	6	6360	6536	gi 1655715	Bztd [Rhodobacter capsulatus]	75	55	177
83	6	1938	2975	gnl PID e323529	putative Plax protein [Bacillus subtilis]	75	56	1038
93	11	7368	5317	gi 39989	methionyl-tRNA synthetase [Bacillus stearothermophilus]	75	58	2052
93	13	9409	8699	gi 1591493	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	75	54	711
95	1	1795	47	gnl PID e323510	Ylov protein [Bacillus subtilis]	75	57	1749
103	2	362	1186	gnl PID e266928	unknown [Mycobacterium tuberculosis]	75	64	825
104	1	691	915	gi 460026	repressor protein [Streptococcus pneumoniae]	75	54	225
113	5	2951	3883	gnl PID d101119	ABC transporter subunit [Synecocystis sp.]	75	55	933
121	1	320	1390	gi 2145131	repressor of class I heat shock gene expression HrcA [Streptococcus mutans]	75	58	1071
127	6	2614	3000	gi 1500451	M. jannaschii predicted coding region MJ1558 [Methanococcus jannaschii]	75	44	387
137	18	10082	10687	gi 393116	P-glycoprotein 5 [Entamoeba histolytica]	75	52	606
149	11	8499	9338	gnl PID d100582	unknown [Bacillus subtilis]	75	55	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
151	6	9100	7673	gi 40467	Hds polypeptide, part of CfrA family [Citrobacter freundii]	75	57	1428
158	1	986	3	gnl PID e253891	UDP-glucose 4-epimerase [Bacillus subtilis]	75	63	984
172	8	5653	6774	gi 142978	glycerol dehydrogenase [Bacillus stearothermophilus]	75	56	1122
172	9	7139	9730	gnl PID e268456	unknown [Mycobacterium tuberculosis]	75	58	2592
173	1	261	79	gnl PID e236469	C10C5.6 [Caenorhabditis elegans]	75	50	183
185	3	3066	2014	gi 1574806	spermidine/putrescine transport ATP-binding protein (potA) [Haemophilus influenzae]	75	56	1053
191	6	5235	4213	gi 149518	phosphoribosyl anthranilate transferase [Lactococcus lactis]	75	61	1023
226	2	1774	1181	gi 2314588	[AE000642] conserved hypothetical protein [Helicobacter pylori]	75	65	594
231	1	1	153	gi 40173	homolog of E.coli ribosomal protein L21 [Bacillus subtilis]	75	57	153
234	1	2	418	gi 2293359	[AF008220] Yqf [Bacillus subtilis]	75	59	417
279	1	552	151	gi 1119198	unknown protein [Bacillus subtilis]	75	50	402
291	7	3558	3827	gi 40011	ORF17 (AA 1-161) [Bacillus subtilis]	75	48	270
375	2	137	628	gi 410137	ORFX13 [Bacillus subtilis]	75	58	492
6	20	16721	17560	gi 2293323	[AF008220] YcdI [Bacillus subtilis]	74	53	840
7	6	4682	6052	gi 1354211	PET12-like protein [Bacillus subtilis]	74	60	1371
18	4	3341	2427	gnl PID d101319	YqgI [Bacillus subtilis]	74	54	915
21	6	5885	4800	gi 1072381	glutaryl-aminopeptidase [Lactococcus lactis]	74	59	1086
24	2	739	548	gi 2314762	[AE000655] ABC transporter, permease protein (yaeE) [Helicobacter pylori]	74	46	192
25	1	2	367	gnl PID d100932	H2O-forming NADH Oxidase [Streptococcus mutans]	74	63	366
38	18	11432	12964	gi 537034	ORF_0488 [Escherichia coli]	74	57	1533
48	10	8924	6669	gi 1513069	P-type adenosine triphosphatase [Listeria monocytogenes]	74	53	2256
55	11	11964	11401	gnl PID e283110	femD [Staphylococcus aureus]	74	64	564
61	2	1782	427	gi 2293216	[AF008220] putative UDP-N-acetylmuramate-alanine ligase [Bacillus subtilis]	74	55	1356
76	10	9414	8065	gnl PID d101325	YqIB [Bacillus subtilis]	74	54	1350
83	2	666	926	pir C33496 C334	hisc homolog - Bacillus subtilis	74	55	261
86	9	8985	8080	gi 683585	prephenate dehydratase [Lactococcus lactis]	74	55	906

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
102	5	5005	5652	gi1143394	OMP-PRPP transferase [Bacillus subtilis]	74	57	648
103	5	4364	3267	gnl PID e323524	Y10N protein [Bacillus subtilis]	74	62	1098
108	7	6864	7592	gnl PID e257631	methyltransferase [Lactococcus lactis]	74	56	729
131	2	478	146	gnl PID d101320	Yqg2 [Bacillus subtilis]	74	45	333
133	2	1380	919	gnl PID e313025	hypothetical protein [Bacillus subtilis]	74	60	462
137	9	6167	6787	gnl PID d100479	Na+ -ATPase subunit D [Enterococcus hirae]	74	53	621
149	4	3008	3883	gnl PID d100581	high level kasamycin resistance [Bacillus subtilis]	74	55	876
157	2	243	824	gi11573373	methylated-DNA--protein-cysteine methyltransferase (dat1) [Haemophilus influenzae]	74	48	582
164	6	3515	4249	gi1410131	OREX7 [Bacillus subtilis]	74	48	735
167	7	5446	5201	gi1413927	ipa-3r gene product [Bacillus subtilis]	74	55	246
171	1	1	1818	gnl PID d102251	beta-galactosidase [Bacillus circulans]	74	62	1818
172	4	1064	2392	gi1466474	cellobiose phosphotransferase enzyme II'' [Bacillus stearothermophilus]	74	50	1329
185	1	326	3	gi11573646	Mg(2+) transport ATPase protein C (mgTC) (SP.P22037) [Haemophilus influenzae]	74	68	324
188	2	1089	2018	gi11573008	ATP dependent translocator homolog (mba) [Haemophilus influenzae]	74	44	930
189	11	6491	7174	gi11661199	sakacin A production response regulator [Streptococcus mutans]	74	60	684
210	2	520	1287	gi12293207	(AF008220) YtmQ [Bacillus subtilis]	74	60	768
261	1	836	192	gi1666983	putative ATP binding subunit [Bacillus subtilis]	74	55	645
263	3	1619	3655	gi1663232	Similarity with S. cerevisiae hypothetical 137.7 kD protein in subtelomeric Y' repeat region [Saccharomyces cerevisiae]	74	42	2017
265	2	844	1227	gi149272	Asparaginase [Bacillus licheniformis]	74	64	384
368	1	1	942	gi1603998	unknown [Saccharomyces cerevisiae]	74	39	942
7	16	13357	11921	gnl PID d101324	YqhX [Bacillus subtilis]	73	57	1437
17	10	5706	5449	gnl PID e305362	unnamed protein product [Streptococcus thermophilus]	73	47	258
31	2	522	244	gnl PID d100576	single strand DNA binding protein [Bacillus subtilis]	73	55	279
32	6	5667	6194	gnl PID d101315	YqfG [Bacillus subtilis]	73	58	528
34	15	10281	9790	gnl PID d102151	(AB001684) ORP42c [Chlorella vulgaris]	73	46	492

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
40	12	9876	9226	gi 1173517	riboflavin synthase alpha subunit (Actinobacillus pleuropneumoniae)	73	55	651
55	2	3592	839	gnl PID d101887	cation-transporting ATPase PacL (Synechocystis sp.)	73	60	2754
55	18	17494	16586	gnl PID e265580	unknown (Mycobacterium tuberculosis)	73	52	909
65	16	7213	7767	gi 143419	ribosomal protein L6 (Bacillus stearothermophilus)	73	60	555
66	3	3300	3659	gnl PID e269883	LacF (Lactobacillus casei)	73	52	360
70	10	5557	5733	gi 1857631	envelope protein (Human immunodeficiency virus type 1)	73	60	177
71	4	6133	8262	gnl PID e322063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	73	45	2130
72	1	3	851	gi 2293177	(AF008220) transporter (Bacillus subtilis)	73	50	849
76	7	7019	6195	gnl PID d101325	YqjF (Bacillus subtilis)	73	66	825
76	12	10009	9533	gi 1573086	uridine kinase (uridine monophosphokinase) (udk) (Haemophilus influenzae)	73	54	477
80	7	8113	9372	gi 1377823	aminopeptidase (Bacillus subtilis)	73	60	1260
97	5	3389	1668	gnl PID d101954	dihydroxyacid dehydratase (Synechocystis sp.)	73	54	1722
98	9	6912	7619	gnl PID e314991	FtsE (Mycobacterium tuberculosis)	73	54	708
108	11	10928	10440	gi 1388109	regulatory protein (Enterococcus faecalis)	73	54	489
128	6	3632	4222	gi 1685111	orf1091 (Streptococcus thermophilus)	73	63	591
138	2	1575	394	gi 147326	transport protein (Escherichia coli)	73	60	1182
140	13	12538	11903	gi E53402 E534	serine O-acetyltransferase (EC 2.3.1.30) - Bacillus stearothermophilus	73	55	636
162	5	5701	4991	gnl PID e223511	putative YhaQ protein (Bacillus subtilis)	73	50	711
164	4	2323	2790	gi 1592076	hypothetical protein (SF:P25768) (Methanococcus jannaschii)	73	52	468
164	8	4815	5546	gi 410137	ORFX13 (Bacillus subtilis)	73	56	732
170	5	4394	5302	gnl PID d100959	homologue of unidentified protein of E. coli (Bacillus subtilis)	73	46	909
178	7	3893	4855	gi 46242	modulation protein B, 5'-end (Rhizobium loti)	73	56	963
204	6	5086	4278	gnl PID e214719	picR protein (Bacillus thuringiensis)	73	41	819
213	2	812	2037	gi 1565296	ribosomal protein S1 homologue; sequence specific DNA-binding protein (Leuconostoc lactis)	73	55	1206
231	2	84	287	gi 40173	homolog of E.coli ribosomal protein L21 (Bacillus subtilis)	73	61	204
237	1	2	505	gi 1773151	adenine phosphoribosyltransferase (Escherichia coli)	73	51	504

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
269	1	2	691	gnl pid d101328	YqjX [Bacillus subtilis]	73	36	690
289	2	1272	832	plr a02771 r7mc	ribosomal protein L7/L12 - Micrococcus luteus	73	66	441
343	1	14	484	gi 1788125	(AE000276) hypothetical 30.4 kD protein in man2-cspC intergenic region [Escherichia coli]	73	47	471
356	1	222	4	gi 2149905	D-glutamic acid adding enzyme (Enterococcus faecalis)	73	50	219
7	5	3165	4691	gnl pid d101833	amidase [Synecocystis sp.]	72	52	1527
7	9	7195	7647	gi 146976	nusB [Escherichia coli]	72	54	453
7	17	13743	13300	gnl pid e289141	similar to hydroxymyristoyl-(acyl carrier protein) dehydratase (Bacillus subtilis)	72	59	444
22	19	15637	16224	gnl pid d101929	ribosome releasing factor [Synecocystis sp.]	72	51	588
33	17	12111	11425	gnl pid d101190	ORF3 [Streptococcus mutans]	72	55	687
34	7	7147	5627	gi 396501	aspartyl-tRNA synthetase [Thermus thermophilus]	72	52	1521
38	23	15372	16085	plr h64108 h641	L-ribulose-phosphate 4-epimerase (araD) homolog - Haemophilus influenzae (strain Rd KW20)	72	54	714
39	5	5094	6905	gnl pid e254877	unknown [Mycobacterium tuberculosis]	72	56	1812
40	6	4469	4636	gi 153672	lactose repressor [Streptococcus mutans]	72	58	168
48	2	1459	1253	gi 310380	Inhibin beta-A-subunit [Ovis aries]	72	33	207
48	29	21729	22424	gi 2314329	(AE000623) glutamine ABC transporter, permease protein (glnP) [Helicobacter pylori]	72	49	696
50	5	4529	3288	gi 1750108	YnbA [Bacillus subtilis]	72	54	1242
51	3	1044	2282	gi 2293230	(AF008220) YtbJ [Bacillus subtilis]	72	54	1239
52	13	13681	13938	gi 142521	deoxyribodipyrimidine photolyase [Bacillus subtilis]	72	45	258
55	1	841	35	gi 882518	ORF_o304, GTG start [Escherichia coli]	72	59	807
75	5	2832	3191	gnl pid e209886	mercuric resistance operon regulatory protein [Bacillus subtilis]	72	44	360
76	6	6229	5771	gi 142450	YhrC protein [Bacillus subtilis]	72	53	459
79	5	5065	4592	gi 2293279	(AF008220) YtcG [Bacillus subtilis]	72	46	474
87	14	14726	12309	gnl pid e23502	putative P1A protein [Bacillus subtilis]	72	52	2418
91	1	444	662	gi 500691	YWO1 gene product [Saccharomyces cerevisiae]	72	50	219
91	7	4516	4764	gi 829615	skeletal muscle sodium channel alpha-subunit [Equus caballus]	72	38	249

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
95	2	2004	1717	gnl pid e33527	putative Asp23 protein (Bacillus subtilis)	72	40	288
109	1	1452	118	gi 143331	alkaline phosphatase regulatory protein (Bacillus subtilis)	72	52	1335
126	1	3	2192	gnl pid d101831	glutamine-binding periplasmic protein (Synechocystis sp.)	72	46	2190
130	3	1735	2478	gi 2415396	(AF015775) carboxypeptidase (Bacillus subtilis)	72	53	744
137	6	2585	2929	gi 472922	v-type Na-ATPase (Enterococcus hirae)	72	46	345
140	10	9601	9203	gi 49224	URF 4 (Synechococcus sp.)	72	48	399
146	5	1906	1247	gnl pid e324945	hypothetical protein (Bacillus subtilis)	72	45	660
147	2	2084	1083	gnl pid e325016	hypothetical protein (Bacillus subtilis)	72	56	1002
147	5	6156	5146	gi 472327	TPP-dependent acetoin dehydrogenase beta-subunit (Clostridium magnum)	72	56	1011
148	8	5381	6433	gi 974332	NAD(P)H-dependent dihydroxyacetone-phosphate reductase (Bacillus subtilis)	72	54	1053
148	14	10256	9675	gnl pid d101319	YqgN (Bacillus subtilis)	72	50	582
159	8	4005	4949	gi 1788770	(AF000330) 0463: 24 pct identical (44 gaps) to 338 residues from penicillin-binding protein 4', PAPB-BACSU SW: p32959 (451 aa) (Escherichia coli)	72	43	945
172	10	9907	10620	gi 763387	unknown (Saccharomyces cerevisiae)	72	55	714
220	3	2862	3602	gi 1574175	hypothetical (Haemophilus influenzae)	72	50	741
267	1	3	449	gi 290513	f470 (Escherichia coli)	72	48	447
281	2	899	540	gnl pid d100964	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis (Bacillus subtilis)	72	45	360
290	1	1018	14	gi 474195	This ORF is homologous to a 40.0 kd hypothetical protein in the htrB 3' region from E. coli, Accession Number X61000 (Mycoplasma-like organism)	72	54	1005
300	1	63	587	gi 746399	transcription elongation factor (Escherichia coli)	72	50	525
316	1	1326	4	gi 158127	protein kinase C (Drosophila melanogaster)	72	40	1323
342	1	227	3	gnl pid d101164	unknown (Bacillus subtilis)	72	54	225
354	1	1	1005	gnl pid d102048	C. thermocellum beta-glucosidase; p26208 (985) (Bacillus subtilis)	72	52	1005
6	10	8134	10467	gnl pid e264229	unknown (Mycobacterium tuberculosis)	71	57	2334
7	20	16231	15464	gi 18046	3-oxoacyl-(acyl-carrier protein) reductase (Cuphea lanceolata)	71	52	768
15	1	1297	2	gnl pid d100571	replicative DNA helicase (Bacillus subtilis)	71	51	1296
15	4	4435	3869	gi 499384	orf189 (Bacillus subtilis)	71	47	567

TABLE 2
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
18	6	5120	4218	gnl PID d101318	YggG [Bacillus subtilis]	71	51	903
29	1	1	540	gi 1773142	similar to the 20.2kd protein in TETB-EXOA region of B. subtilis [Escherichia coli]	71	56	540
38	20	13327	13830	gi 537036	ORF_0158 [Escherichia coli]	71	48	504
51	12	15015	12676	gi 149528	dipeptidyl peptidase IV [Lactococcus lactis]	71	55	2340
55	23	21040	20585	gi 2343285	(AF015453) surface located protein [Lactobacillus rhamnosus]	71	58	456
60	2	705	265	gnl PID d101320	Ygg2 [Bacillus subtilis]	71	44	441
71	18	24679	26226	gi 580920	rodD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis]	71	44	1548
71	25	30587	30360	gi 606028	ORF_0414; Geneplot suggests frameshift near start but none found [Escherichia coli]	71	50	228
72	6	5239	6729	gi 580835	lysine decarboxylase [Bacillus subtilis]	71	48	1491
72	14	11191	12878	gi 624085	similar to rat beta-alanine synthetase encoded by GenBank Accession Number 527881; contains ATP/GTP binding motif [Paramecium bursaria Chlorella virus 1]	71	54	888
73	11	7269	7033	gi 1906594	Pnl [Rattus norvegicus]	71	42	237
74	6	10385	8517	gi 1573733	prolyl-tRNA synthetase (proS) [Haemophilus influenzae]	71	52	1869
81	9	5772	6578	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	71	45	807
86	5	4602	3604	gnl PID e322063	ss-1,4-galactosyltransferase [Streptococcus pneumoniae]	71	53	999
105	4	3619	4707	gi 2323341	(AF014460) PepQ [Streptococcus mutans]	71	58	1089
106	13	13557	12955	gi 1519287	LemA [Listeria monocytogenes]	71	48	603
114	2	1029	1979	gi 310303	mosA [Rhizobium meliloti]	71	55	951
122	2	564	1205	gi 1649037	glutamine transport ATP-binding protein GLNQ [Salmonella typhimurium]	71	50	642
132	5	9018	7063	gnl PID d102049	H. influenzae hypothetical ABC transporter; P44808 (974) [Bacillus subtilis]	71	51	1956
140	1	1141	227	gi 1673788	(AB0000015) Mycoplasma pneumoniae, fructose-bisphosphate aldolase; similar to Swiss-Prot Accession Number P13243, from B. subtilis [Mycoplasma pneumoniae]	71	49	915
140	5	5635	4973	gnl PID d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus [Bacillus subtilis]	71	48	663
141	7	7369	7845	gnl PID d102005	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN E. COLI AND MYCOPLASMA PNEUMONIAE. [Bacillus subtilis]	71	51	477

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
193	1	1	165	gi 46912	ribosomal protein L13 (Staphylococcus carnosus)	71	59	165
194	3	2205	1594	gi 535351	CodY (Bacillus subtilis)	71	52	612
199	3	1510	1319	gi 2182574	(AE000090) Y4pE (Rhizobium sp. NGR234)	71	45	192
208	2	2616	3752	gi 11787378	(AE000213) hypothetical protein in purB 5' region (Escherichia coli)	71	57	1137
209	2	2022	1141	gi 41432	fepC gene product (Escherichia coli)	71	46	882
210	5	1911	3071	gi 49316	ORF2 gene product (Bacillus subtilis)	71	45	1161
210	6	3069	3386	gi 580900	ORF3 gene product (Bacillus subtilis)	71	48	318
212	2	3561	1381	gi 557567	ribonucleotide reductase R1 subunit (Mycobacterium tuberculosis)	71	53	2181
233	3	2003	2920	gnl PID d101320	Yqqr (Bacillus subtilis)	71	50	918
244	1	13	1053	gnl PID d100964	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis (Bacillus subtilis)	71	55	1041
251	2	1008	1874	gi 755601	unknown (Bacillus subtilis)	71	46	867
282	2	906	712	gi 11353874	unknown (Rhodobacter capsulatus)	71	46	195
312	4	2137	1565	gnl PID d102245	(AB005554) yxbF (Bacillus subtilis)	71	34	573
338	1	3	683	gi 1591045	hypothetical protein (SP:P31466) (Methanococcus jannaschii)	71	48	681
346	1	3	164	gi 1591234	hypothetical protein (SP:P42297) (Methanococcus jannaschii)	71	36	162
374	1	619	2	gi 397526	clumping factor (Staphylococcus aureus)	71	23	618
377	1	688	2	gi 397526	clumping factor (Staphylococcus aureus)	71	23	687
3	8	7419	6958	gnl PID e269486	unknown (Bacillus subtilis)	70	42	462
3	10	8395	9075	gnl PID e255543	putative iron dependant repressor (Staphylococcus epidermidis)	70	46	681
7	14	11024	10254	gnl PID d100290	undefined open reading frame (Bacillus stearothermophilus)	70	55	771
7	18	14213	13719	gnl PID d101090	biotin carboxyl carrier protein of acetyl-CoA carboxylase (Synechocystis sp.)	70	56	495
9	2	1057	287	gnl PID d100581	unknown (Bacillus subtilis)	70	52	771
12	4	2610	1789	gnl PID d101195	yycJ (Bacillus subtilis)	70	52	822
21	2	2586	1846	gi 2293447	(AF008930) ATPase (Bacillus subtilis)	70	54	741
22	13	10955	11512	gi 1163295	Ydr540cp (Saccharomyces cerevisiae)	70	50	558
30	6	4315	3980	gi 39478	ATP binding protein of transport ATPases (Bacillus firmus)	70	51	336

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
31	1	370	113	gi 662792	single-stranded DNA binding protein [unidentified eubacterium]	70	36	258
33	15	10839	9521	gi 1161219	homologous to D-amino acid dehydrogenase enzyme [Pseudomonas aeruginosa]	70	50	1119
38	6	3812	4312	gi 2058547	ComYD [Streptococcus gordonii]	70	48	501
38	25	17986	18477	gi 537033	ORF f356 [Escherichia coli]	70	58	492
40	13	11054	9846	gi 1173516	riboflavin-specific deaminase [Actinobacillus pleuropneumoniae]	70	52	1209
42	2	722	1954	gi 1146183	putative [Bacillus subtilis]	70	51	1233
43	3	2373	1612	gi 1591493	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	70	48	762
45	8	9197	8049	gnl pid d102036	subunit of ADP-glucose pyrophosphorylase [Bacillus stearothermophilus]	70	54	1149
59	2	567	956	gnl pid d100302	neopullulanase [Bacillus sp.]	70	42	390
60	3	1874	795	gnl pid e276466	aminopeptidase P [Lactococcus lactis]	70	48	1080
61	4	5553	2437	gnl pid e275074	SNF [Bacillus cereus]	70	51	3117
61	7	7914	6802	gi 1573037	cystathionine gamma-synthase (metB) [Haemophilus influenzae]	70	52	1113
63	7	5372	7222	gnl pid d100974	unknown [Bacillus subtilis]	70	54	1851
68	7	7126	6962	gi 1263014	emm18.1 gene product [Streptococcus pyogenes]	70	37	165
72	12	10081	10911	gi 2313093	(AE000524) carboxynorspermidine decarboxylase (nspC) [Helicobacter pylori]	70	56	831
75	10	7888	8124	gi 1877423	galactose-1-P-uridyl transferase [Streptococcus mutans]	70	59	237
79	3	3424	2525	gi 39881	ORF 311 (AA 1-311) [Bacillus subtilis]	70	47	900
87	10	9369	7324	gnl pid e233506	putative Pkn2 protein [Bacillus subtilis]	70	52	2046
96	14	10640	11788	gi 1573209	tRNA-guanine transglycosylase (tgt) [Haemophilus influenzae]	70	52	1149
113	2	574	1086	gi 433630	A180 [Saccharomyces cerevisiae]	70	59	513
123	5	2901	3461	gnl pid d100585	unknown [Bacillus subtilis]	70	45	561
125	5	4593	4282	gnl pid e276674	capacitative calcium entry channel 1 [Bos taurus]	70	35	312
129	5	4500	3454	gnl pid d101314	Yqer [Bacillus subtilis]	70	47	1047
133	3	2608	1394	gi 2293312	(AF008220) Ytcf [Bacillus subtilis]	70	50	1215
135	1	420	662	gnl pid e265530	yorfe [Streptococcus pneumoniae]	70	47	243
137	3	438	932	gi 472919	v-type Na-ATPase [Enterococcus hirae]	70	57	495
138	1	440	3	gi 147336	transmembrane protein [Escherichia coli]	70	42	438

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
140	16	18796	16364	gi 976441	N5-methyltetrahydrofolate homocysteine methyltransferase [Saccharomyces cerevisiae]	70	53	2433
167	10	8263	6695	gi 149535	D-alanine activating enzyme [Lactobacillus casei]	70	52	1569
204	4	3226	2747	gnl PID d102049	E. coli hypothetical protein; P31805 (267) [Bacillus subtilis]	70	51	480
207	3	2627	2869	gnl PID e309213	racGAP [Dictyostellium discoideum]	70	45	243
282	3	1136	882	gi 1353874	unknown [Rhodobacter capsulatus]	70	50	255
6	21	17554	18453	gnl PID e233879	hypothetical protein [Bacillus subtilis]	69	44	900
6	22	18482	19471	gi 580883	lpa-88d gene product [Bacillus subtilis]	69	53	990
22	6	4682	5824	gi 2209379	[AF006720] ProJ [Bacillus subtilis]	69	48	1143
22	9	7992	8651	gnl PID d100580	unknown [Bacillus subtilis]	69	51	660
22	12	9871	10767	gnl PID d100581	unknown [Bacillus subtilis]	69	51	897
27	7	5857	5348	gnl PID d102012	(AB001488) FUNCTION UNKNOWN. [Bacillus subtilis]	69	28	510
36	10	7294	10116	gi 437916	isoleucyl-tRNA synthetase [Staphylococcus aureus]	69	53	2823
38	1	2	1090	gi 141900	alcohol dehydrogenase (EC 1.1.1.1) [Alcaligenes eutrophus]	69	48	1089
40	14	11333	11944	gi 1573280	Holliday junction DNA helicase (ruva) [Haemophilus influenzae]	69	44	612
40	15	11942	12517	gi 1573653	DNA-3-methyladenine glycosidase I (tagI) [Haemophilus influenzae]	69	50	576
45	6	6947	5490	gi 580887	starch (bacterial glycogen) synthase [Bacillus subtilis]	69	47	1458
48	34	24932	24153	gnl PID e233870	hypothetical protein [Bacillus subtilis]	69	36	780
49	6	6183	6521	gi 396297	similar to phosphotransferase system enzyme II [Escherichia coli]	69	50	339
49	8	7586	8338	gi 396420	similar to Alcaligenes eutrophus PHC1 D-ribulose-5-phosphate 3 epimerase [Escherichia coli]	69	49	753
55	6	8262	7033	gi 1146238	poly(A) polymerase [Bacillus subtilis]	69	50	1230
59	3	954	2333	gnl PID e313038	hypothetical protein [Bacillus subtilis]	69	54	1380
62	3	1170	1418	gnl PID d101915	hypothetical protein [Synecocystis sp.]	69	49	249
63	8	7298	7762	gi 293017	ORF3 (put.); putative [Lactococcus lactis]	69	42	465
66	4	3657	5081	gi 153755	phospho-beta-D-galactosidase (EC 3.2.1.85) [Lactococcus lactis cremoris]	69	49	1425
66	5	5126	6829	gi 433809	enzyme II [Streptococcus mutans]	69	46	1704
71	6	10017	10664	gnl PID e322063	ss-1,4-galactosyltransferase [Streptococcus pneumoniae]	69	39	648

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
71	21	27730	27966	gnl PID d100649	DE-cadherin [Drosophila melanogaster]	69	30	237
77	1	1	237	gi 287870	groES gene product [Lactococcus lactis]	69	44	237
81	5	3622	4101	gi 1573605	fucose operon protein (fucU) [Haemophilus influenzae]	69	52	480
83	1	40	714	pir C33496 C334	hisc homolog - Bacillus subtilis	69	46	675
83	16	15742	16335	gi 143372	[phosphoribosyl] glycinamide formyltransferase (PUR-N) [Bacillus subtilis]	69	46	594
85	2	1212	916	gi 194097	IFN-response element binding factor 1 [Mus musculus]	69	48	297
91	5	3678	4274	gi 1574712	anaerobic ribonucleoside-triphosphate reductase activating protein (nrtdg) [Haemophilus influenzae]	69	44	597
98	5	3247	4032	gnl PID d100262	LivP protein [Salmonella typhimurium]	69	51	786
108	5	4085	5056	gnl PID e257629	transcription factor [Lactococcus lactis]	69	49	972
126	3	3078	4568	gnl PID d101329	YqjJ [Bacillus subtilis]	69	49	1491
131	6	4121	2889	gnl PID d101314	Yqer [Bacillus subtilis]	69	47	1233
136	2	1505	2299	gnl PID d100581	unknown [Bacillus subtilis]	69	47	795
149	5	3852	4763	gnl PID e323525	YloQ protein [Bacillus subtilis]	69	50	912
149	12	9316	10655	gi 151571	Homology with E.coli and P.aeruginosa lysA gene; product of unknown function; putative [Pseudomonas syringae]	69	52	1320
153	4	3191	3829	gi 1710373	BrnQ [Bacillus subtilis]	69	44	639
169	3	849	2324	gnl PID d100582	temperature sensitive cell division [Bacillus subtilis]	69	49	1476
180	1	566	3	gi 488339	alpha-amylase [unidentified cloning vector]	69	50	564
212	1	1196	231	gi 1395209	ribonucleotide reductase R2-2 small subunit [Mycobacterium tuberculosis]	69	53	966
226	1	2	661	pir J02285 J022	nodulin-26 - soybean	69	41	660
233	5	3249	4766	gi 472918	v-type Na-ATPase [Enterococcus hirae]	69	56	1518
235	3	660	1766	gi 148945	methylase [Haemophilus influenzae]	69	43	1107
243	2	865	2361	gnl PID d100225	ORF5 [Barley yellow dwarf virus]	69	69	1497
251	3	2899	1967	gi 2289231	macrolide-efflux protein [Streptococcus agalactiae]	69	51	933
310	1	1	282	gnl PID e322442	peptide deformylase [Clostridium beijerinckii]	69	55	282
369	1	868	2	gi 397526	clumping factor [Staphylococcus aureus]	69	22	867
370	1	749	3	gi 397526	clumping factor [Staphylococcus aureus]	69	21	747

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins 'similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
379	1	44	280	gnl pid d100649	DE-cadherin [Drosophila melanogaster]	69	30	237
388	1	260	72	gi 1787524	(AE000225) hypothetical 32.7 kD protein in trpL-bttr intergenic region [Escherichia coli]	69	44	189
1	2	2006	3040	gnl pid d101809	ABC transporter [Synecocystis sp.]	68	43	1035
12	5	3958	2600	gi 2182992	histidine kinase [Lactococcus lactis cremoris]	68	45	1359
15	2	1790	1311	pir s16974 RSBS	ribosomal protein L9 - Bacillus stearothermophilus	68	56	480
16	6	7353	5701	gi 1787041	(AE000184) o510; This 530 aa orf is 33 pct identical (14 gaps) to 525 residues of an approx. 640 aa protein YHES_HAEIN SW: P44808 [Escherichia coli]	68	45	1653
17	12	6479	6805	gi 553165	acetylcholinesterase (Homo sapiens)	68	68	327
20	13	14128	14505	gi 142700	P competence protein (ttg start codon) (put.); putative [Bacillus subtilis]	68	40	378
22	32	24612	25397	gi 289262	comE ORF3 [Bacillus subtilis]	68	36	786
30	7	4548	4288	gi 311388	ORF1 [Azorhizobium caulinodans]	68	46	261
36	5	3911	4585	gi 1573041	hypothetical [Haemophilus influenzae]	68	54	675
46	6	5219	6040	gi 1790131	(AE000446) hypothetical 29.7 kD protein in lbpA-gvrB intergenic region [Escherichia coli]	68	47	822
54	10	6235	7086	gi 882579	CG Site No. 29739 [Escherichia coli]	68	55	852
55	5	7069	5165	gnl pid d101914	ABC transporter [Synecocystis sp.]	68	45	1905
71	3	6134	5613	gi 1573353	outer membrane integrity protein (tolA) [Haemophilus influenzae]	68	50	522
71	10	15342	16613	gi 580866	lpa-12d gene product [Bacillus subtilis]	68	31	1272
71	12	17560	18792	gi 44073	SecY protein [Lactococcus lactis]	68	35	1233
71	17	22295	24703	gi 1762349	Involved in protein export [Bacillus subtilis]	68	50	2409
73	16	10208	9729	gi 1353537	dUTPase [Bacteriophage rlt]	68	51	480
86	18	17198	16011	gi 413943	lpa-19d gene product [Bacillus subtilis]	68	53	1188
87	17	17491	15866	gi 150209	ORF 1 [Mycoplasma mycoides]	68	43	1626
89	6	5139	4354	gi 1498824	M. jannaschii predicted coding region MJ0062 [Methanococcus jannaschii]	68	40	786
89	11	8021	8242	gi 150974	4-oxalocrotonate tautomerase [Pseudomonas putida]	68	43	222
97	8	6755	5394	gi 2367358	(AE000491) hypothetical 52.9 kD protein in ald8-rpsF intergenic region [Escherichia coli]	68	41	1362

TABLE 2
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
98	3	1418	2308	gnl PID d100261	liuA protein (Salmonella typhimurium)	68	40	891
99	13	16414	17280	gi 455363	regulatory protein (Streptococcus mutans)	68	50	867
115	3	5054	3693	gi 466474	cellobiose phosphorylase enzyme II' (Bacillus stearothermophilus)	68	44	1362
124	7	3394	3221	gnl PID d100702	cut14 protein (Schizosaccharomyces pombe)	68	56	174
125	2	2923	1922	gi 450566	transmembrane protein (Bacillus subtilis)	68	50	1002
132	2	4858	2888	gnl PID d101732	DNA ligase (Synecocystis sp.)	68	52	1971
140	7	7765	7580	gi 1209711	unknown (Saccharomyces cerevisiae)	68	47	186
150	1	539	3	gi 402490	ADP-ribosylarginine hydrolase (Mus musculus)	68	59	537
164	1	58	867	gnl PID e25114	glutamate racemase (Bacillus subtilis)	68	49	810
164	2	819	1835	gnl PID e25117	hypothetical protein (Bacillus subtilis)	68	50	1017
169	7	3946	4104	pir B54545 B545	hypothetical protein - Lactococcus lactis subsp. lactis plasmid pSL2	68	40	159
170	4	4247	4396	gi 304146	spore coat protein (Bacillus subtilis)	68	52	150
171	8	6002	7054	gi 38722	precursor (aa -20 to 381) (Acinetobacter calcoaceticus)	68	54	1053
198	3	2473	1871	gnl PID e313075	hypothetical protein (Bacillus subtilis)	68	46	603
211	2	969	1802	gi 1439528	E1IC-man (Lactobacillus curvatus)	68	45	834
214	8	4926	4231	gnl PID d102049	H. influenzae hypothetical protein; P43990 (182) (Bacillus subtilis)	68	50	696
217	6	4955	5170	gnl PID e326966	similar to B. vulgaris CMS-associated mitochondrial ... (reverse transcriptase) (Arabidopsis thaliana)	68	36	216
218	7	3930	4745	gi 2293198	(AF008220) YtgP (Bacillus subtilis)	68	38	816
220	6	4628	4338	gnl PID e325791	(AJ000005) orf1 (Bacillus megaterium)	68	51	291
236	1	746	108	gi 410137	ORFX13 (Bacillus subtilis)	68	46	639
237	2	675	1451	gi 396348	homoserine transuccinylase (Escherichia coli)	68	49	777
250	4	771	1229	gi 310859	ORF2 (Synecococcus sp.)	68	50	459
254	1	517	155	gi 1787105	(AE000189) 0648 was 0669; This 669 aa orf is 40 pct identical (1 gaps) to 217 residues of an approx. 232 aa protein Y8BA_HAEIN SW: P45247 (Escherichia coli)	68	44	363
337	1	1	774	gnl PID e261990	putative orf (Bacillus subtilis)	68	47	774
345	1	3	653	gi 149513	thymidylate synthase (EC 2.1.1.45) (Lactococcus lactis)	68	61	651

TABLE 2
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
386	2	417	4	gi11573353	outer membrane integrity protein (tolA) (Haemophilus influenzae)	68	51	414
2	4	5722	4697	gi11592141	M. jannaschii predicted coding region MJ1507 (Methanococcus jannaschii)	67	26	1026
3	6	5397	4591	gi12993175	(AF008220) signal transduction regulator (Bacillus subtilis)	67	44	807
5	2	2301	574	gi12113365	(AE000547) para-aminobenzoate synthetase (pabB) (Helicobacter pylori)	67	48	1728
6	19	16063	16758	gi1413931	lipa-7d gene product (Bacillus subtilis)	67	41	696
22	8	7094	7897	gi11928962	pyroline-5-carboxylate reductase (Actinidia delictosa)	67	51	804
29	10	8335	9072	gi1468745	gtrC gene product (Bacillus brevis)	67	41	738
31	3	1379	585	gi12425123	(AF019986) PKB (Dictyostelium discoideum)	67	49	795
32	11	8849	10150	gi142029	ORF1 gene product (Escherichia coli)	67	47	1302
36	16	14830	15546	gi11592142	ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)	67	43	717
38	9	4958	5392	gnl PID e214803	T22B3.3 (Caenorhabditis elegans)	67	47	435
38	21	13775	14512	gi1537037	ORF_o216 (Escherichia coli)	67	52	738
45	9	10428	9181	gi1551710	branching enzyme (glgB) (EC 2.4.1.18) (Bacillus stearothermophilus)	67	51	1248
48	23	18344	17514	gi1413949	lipa-25d gene product (Bacillus subtilis)	67	50	831
50	2	1773	952	gnl PID d101330	YQJQ (Bacillus subtilis)	67	55	822
53	1	431	3	gi11574291	flabrial transcription regulation repressor (pilB) (Haemophilus influenzae)	67	40	429
55	13	12740	11946	gnl PID e252990	ORF_YDL037c (Saccharomyces cerevisiae)	67	51	795
61	9	9210	8329	gnl PID e264711	ATP-binding cassette transporter A (Staphylococcus aureus)	67	50	882
71	2	5614	6117	gi1197667	vitellogenin (Anolis pulchellus)	67	36	504
81	7	4489	4983	gi1142714	phosphoenolpyruvate:mannose phosphotransferase element IIB (Lactobacillus curvatus)	67	42	495
83	7	2957	3214	gi11276746	Acyl carrier protein (Porphyra purpurea)	67	37	258
86	8	8140	6809	gi11147744	PSR (Enterococcus hirae)	67	45	1332
97	3	986	1366	gnl PID d102235	(AB006631) unnamed protein product (Streptococcus mutans)	67	43	381
102	1	601	1413	gi1682765	mccB gene product (Escherichia coli)	67	36	813
106	3	1109	1987	gi1148921	LicD protein (Haemophilus influenzae)	67	43	879
115	4	5982	5656	gi1895750	putative cellobiose phosphotransferase enzyme III (Bacillus subtilis)	67	44	327

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
115	7	8421	8077	gi 466473	cellobiose phosphotransferase enzyme II' (Bacillus stearothermophilus)	67	51	345
127	13	8127	7021	gi 147326	transport protein (Escherichia coli)	67	45	1107
136	3	2215	2859	gnl PID d100581	unknown (Bacillus subtilis)	67	49	645
140	21	23317	20906	gnl PID d101912	phenylalanyl-tRNA synthetase (Synachocystis sp.)	67	43	2412
146	6	2894	1893	gi 2182994	histidine kinase (Lactococcus lactis cremoris)	67	44	1002
151	8	11476	11117	gnl PID d100085	ORF129 (Bacillus cereus)	67	48	360
160	10	7453	8646	gi 2281317	OrfB: similar to a Streptococcus pneumoniae putative membrane protein encoded by GenBank Accession Number X59400; inactivation of the OrfB gene leads to UV-sensitivity and to decrease of homologous recombination (plasmidic test) (Lactococcus 1)	67	46	1194
163	3	3099	4505	gnl PID d101317	YqfR (Bacillus subtilis)	67	47	1407
167	8	6704	5454	gi 1161933	DitB (Lactobacillus casei)	67	45	1251
169	4	2322	2879	gnl PID d101331	YqkG (Bacillus subtilis)	67	41	558
171	11	7656	8384	gi 153841	pneumococcal surface protein A (Streptococcus pneumoniae)	67	50	729
188	3	1930	3723	gi 1542975	AbcB (Thermotogaobacterium thermosulfurigenes)	67	46	1794
189	6	3599	3141	gnl PID e325178	Hypothetical protein (Bacillus subtilis)	67	52	459
205	3	1663	2211	gi 606073	ORF_0169 (Escherichia coli)	67	47	549
207	4	2896	3456	gi 2276374	DtxR/iron regulated lipoprotein precursor (Corynebacterium diphtheriae)	67	49	561
217	3	4086	3703	gi 893750	putative cellobiose phosphotransferase enzyme III (Bacillus subtilis)	67	42	384
246	2	291	662	gi 1842438	unknown (Bacillus subtilis)	67	43	372
252	1	2	745	gi 2351768	PspA (Streptococcus pneumoniae)	67	41	744
265	3	1134	1811	gi 2313847	(AE000585) L-asparaginase II (ansB) (Helicobacter pylori)	67	42	678
295	1	1	375	gi 2276374	DtxR/iron regulated lipoprotein precursor (Corynebacterium diphtheriae)	67	43	375
1	7	4898	5146	gnl PID e255179	unknown (Mycobacterium tuberculosis)	66	56	249
3	1	389	3	gnl PID e269548	Unknown (Bacillus subtilis)	66	48	387
3	20	19267	20805	gi 39956	TrgC (Bacillus subtilis)	66	50	1539
4	3	2545	2718	gi 11787564	(AE000228) phage shock protein C (Escherichia coli)	66	36	174
5	9	13197	12592	gi 1574291	fimbrial transcription regulation repressor (pilB) (Haemophilus influenzae)	66	46	606

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
9	4	2872	1451	gnl PID e26928	unknown [Mycobacterium tuberculosis]	66	43	1422
12	2	1469	1200	gi 520407	orf2; G7G start codon [Bacillus thuringiensis]	66	42	270
15	12	10979	9897	gi 2314738	[AE000653] translation elongation factor EF-Ts (tsf) [Helicobacter pylori]	66	49	1083
16	2	1312	734	gnl PID d102245	[AB005554] yxbF [Bacillus subtilis]	66	35	579
22	3	1372	1851	gi 1480916	signal peptidase type II [Lactococcus lactis]	66	38	480
22	7	5828	7096	gnl PID e206261	gamma-glutamyl phosphate reductase [Streptococcus thermophilus]	66	51	1269
22	20	16194	17138	gnl PID e281914	yitL [Bacillus subtilis]	66	50	945
30	2	530	976	gi 2314379	[AE000627] ABC transporter, ATP-binding protein (yhcg) [Helicobacter pylori]	66	40	447
32	1	199	984	gi 312444	ORF2 [Bacillus caldolyticus]	66	49	786
33	13	8352	7234	gi 1387979	44% identity over 302 residues with hypothetical protein from <i>Synechocystis</i> sp. accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]	66	44	1119
34	6	5658	4708	gnl PID e250724	orf2 [Lactobacillus sake]	66	39	951
34	14	9792	9574	gi 1590997	M. jannaschii predicted coding region MJ0272 [Methanococcus jannaschii]	66	48	219
35	16	15163	14501	gi 1773352	Cap5M [Staphylococcus aureus]	66	46	663
36	9	6173	6976	gi 1518680	minicell-associated protein DivIVA [Bacillus subtilis]	66	35	804
36	11	10396	10824	bbs 155344	insulin activator factor, INSAP [human, Pancreatic insulinoma, Peptide Partial, 744 aa] [Homo sapiens]	66	43	429
48	1	28	1419	gnl PID e325204	hypothetical protein [Bacillus subtilis]	66	50	1392
48	7	3810	4112	gi 2182574	[AE000090] Y4pE [Rhizobium sp. NGR234]	66	40	303
52	4	3595	2789	gi 388565	major cell-binding factor [Campylobacter jejuni]	66	52	807
54	3	2662	1076	gnl PID d101831	glutamine-binding periplasmic protein [Synechocystis sp.]	66	43	1587
61	10	9740	9183	gnl PID o154144	mdr gene product [Staphylococcus aureus]	66	44	558
72	13	10893	11993	gi 2313129	[AE000526] H. pylori predicted coding region HP0049 [Helicobacter pylori]	66	44	1101
74	9	13267	12476	gi 1573941	hypothetical [Haemophilus influenzae]	66	43	792
75	1	2	868	gi 1574631	nicotinamide mononucleotide transporter (pnuC) [Haemophilus influenzae]	66	48	867
75	7	5303	4275	gi 41312	put. EBG repressor protein [Escherichia coli]	66	40	1029

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
82	7	6813	8123	[gnl PID a255128]	triglycer factor [Bacillus subtilis]	66	53	1311
83	3	905	1219	[pir C33496 C334]	hisc homolog - Bacillus subtilis	66	44	315
86	10	9407	8925	[g1 683584]	shikimate kinase [Lactococcus lactis]	66	41	483
88	10	7001	6060	[gi 2098719]	putative fibrillar-associated protein (Actinomyces naeslundii)	66	52	942
89	1	951	4	[g1 410118]	ORFX19 [Bacillus subtilis]	66	41	948
93	7	3661	2711	[g1 1787936]	{AE000260} f298; This 298 aa orf is 51 pct identical (5 gaps) to 297 residues of an approx. 304 aa protein YCSN_BACSU SW: R42972 (Escherichia coli)	66	49	951
104	3	1805	3049	[g1 1469784]	putative cell division protein ftsW [Enterococcus hirae]	66	48	1245
106	14	11576	14253	[g1 40027]	homologous to E.coli gidB [Bacillus subtilis]	66	52	678
107	3	965	1864	[g1 144858]	ORF A [Clostridium perfringens]	66	49	900
112	7	5718	6593	[g1 609332]	DprA [Haemophilus influenzae]	66	43	876
115	1	3	302	[g1 727367]	Hyrlp [Saccharomyces cerevisiae]	66	56	300
122	1	3	566	[gnl PID d101328]	Yqiy [Bacillus subtilis]	66	36	564
126	8	11759	11046	[gnl PID d101163]	ORF3 [Bacillus subtilis]	66	48	714
128	11	8201	8431	[g1 726288]	growth associated protein GAP-43 [Xenopus laevis]	66	41	231
131	8	4894	4508	[g1 486661]	TMna related protein [Saccharomyces cerevisiae]	66	39	387
140	3	3236	2574	[g1 40056]	phop gene product [Bacillus subtilis]	66	36	663
140	15	16318	15434	[g1 1658189]	[5,10-methylenetetrahydrofolate reductase [Erwinia carotovora]	66	48	885
146	12	7926	7636	[gnl PID d101140]	transposase [Synechocystis sp.]	66	42	291
147	6	7137	6154	[g1 472326]	TPP-dependent acetoin dehydrogenase alpha-subunit [Clostridium magnum]	66	48	984
149	6	4435	5430	[gnl PID d101087]	pentose-5-phosphate-3-epimerase [Synechocystis sp.]	66	46	996
149	13	10754	11575	[g1 42371]	pyruvate formate-lyase activating enzyme (AA 1-246) [Escherichia coli]	66	42	822
186	4	2578	2270	[gnl PID d101199]	ORF11 [Enterococcus faecalis]	66	41	309
207	2	2340	2597	[gnl PID e321893]	envelope glycoprotein gp160 [Human immunodeficiency virus type 1]	66	46	258
210	7	3358	3678	[g1 49318]	ORF4 gene product [Bacillus subtilis]	66	46	331
217	8	5143	5355	[g1 49538]	thrombin receptor [Cricetulus longicaudatus]	66	38	213
220	4	3875	3642	[g1 466648]	alternate name ORF1 of L23635 [Escherichia coli]	66	33	234

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
223	1	1070	138	gnl pid e247187	zinc finger protein [Bacteriophage phigle]	66	45	933
224	2	1864	2640	gi 1176399	putative ABC transporter subunit [Staphylococcus epidermidis]	66	41	777
243	1	3	872	dbj AB000617.2	YcdH [Bacillus subtilis]	66	45	870
268	2	891	568	gi 517210	putative transposase [Streptococcus pyogenes]	66	60	324
322	1	2	643	gi 1499836	2n protease [Methanococcus jannaschii]	66	40	642
5	10	13909	13178	gi 1574292	hypothetical [Haemophilus influenzae]	65	34	732
6	11	10465	11190	gi 142854	homologous to E. coli radC gene product and to unidentified protein from Staphylococcus aureus [Bacillus subtilis]	65	48	726
7	2	647	405	pir C64146 C641	hypothetical protein HI0259 - Haemophilus influenzae (strain Rd KW20)	65	42	243
7	7	6246	6821	gnl pid d101323	YqhU [Bacillus subtilis]	65	50	576
10	2	1873	1397	gi 1163111	ORF-1 [Streptococcus pneumoniae]	65	54	477
16	3	1428	2222	gnl pid e325010	hypothetical protein [Bacillus subtilis]	65	45	795
21	4	3815	3357	gnl pid e314910	hypothetical protein [Staphylococcus sciuri]	65	40	459
22	34	25776	26384	gi 1123030	CpxA [Actinobacillus pleuropneumoniae]	65	42	609
43	2	1648	290	gi 1044826	F14E5.1 [Caenorhabditis elegans]	65	38	1359
48	13	10062	10856	gi 1573390	hypothetical [Haemophilus influenzae]	65	45	795
48	22	17521	16883	gi 1573391	hypothetical [Haemophilus influenzae]	65	37	639
48	25	19027	18533	gnl pid e264484	YCR020c, len:215 [Saccharomyces cerevisiae]	65	38	495
49	3	3856	5334	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	65	32	1479
50	6	5337	4519	gi 171963	tRNA isopentenyl transferase [Saccharomyces cerevisiae]	65	42	819
52	15	14728	15588	gi 1499745	M. jannaschii predicted coding region MJ0912 [Methanococcus jannaschii]	65	46	861
59	7	3963	4745	gi 496514	orf zeta [Streptococcus pyogenes]	65	42	783
68	3	2500	3483	gi 887824	ORF_o310 [Escherichia coli]	65	46	984
69	3	2171	1077	gnl pid e311453	unknown [Bacillus subtilis]	65	42	1095
69	7	6029	5325	gi 809660	deoxyribose-phosphate aldolase [Bacillus subtilis]	65	55	705
71	5	8536	9783	gi 1573324	glycosyl transferase lgtC (GP:U14554.4) [Haemophilus influenzae]	65	42	1248
72	8	7664	8527	gnl pid e267589	Unknown, highly similar to several spermidine synthases [Bacillus subtilis]	65	39	864

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
76	5	5773	4097	gnl PID d101723	DNA REPAIR PROTEIN REC (RECOMBINATION PROTEIN N). [Escherichia coli]	65	44	1677
76	9	8099	7875	gi 1574276	exodeoxyribonuclease, small subunit (xseB) [Haemophilus influenzae]	65	38	225
84	2	2870	2352	gi 2313188	(AE000532) conserved hypothetical protein [Helicobacter pylori]	65	41	519
86	15	14495	13407	gnl PID d101880	3-dehydroquininate synthase [Synecocystis sp.]	65	44	1089
87	3	3706	2423	gi 151239	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas nevaloni]	65	51	1284
88	3	2425	2736	gi 1098510	unknown [Lactococcus lactis]	65	30	312
89	2	1627	1007	gnl PID d102008	(AB001488) SIMILAR TO ORF14 OF ENTEROCOCCUS FAECALIS TRANSPOSON TN916. [Bacillus subtilis]	65	41	621
111	6	6635	6186	gnl PID e246063	NM23/nucleoside diphosphate kinase [Xenopus laevis]	65	50	450
116	1	3	1016	gnl PID d101125	queuosine biosynthesis protein QueA [Synecocystis sp.]	65	44	1014
123	1	69	389	gi 498839	ORF2 [Clostridium perfringens]	65	36	321
123	7	6522	7190	gi 1575577	DNA-binding response regulator [Thermotoga maritima]	65	39	669
125	3	3821	2859	gnl PID e257609	sugar-binding transport protein [Anaerocellum thermophilum]	65	47	963
137	12	8015	7818	gi 2182574	(AE000090) Y4pE [Rhizobium sp. NGR234]	65	41	198
147	4	5031	3885	gi 472329	dihydrolipamide acetyltransferase [Clostridium magnum]	65	47	1137
148	2	1053	1931	gnl PID d101319	YqgH [Bacillus subtilis]	65	42	879
151	2	3212	4687	gi 304897	EcoE type 1 restriction modification enzyme M subunit [Escherichia coli]	65	50	1476
156	2	730	437	gi 310893	membrane protein [Thellieria parva]	65	47	294
164	7	4256	4837	gi 410132	ORFX8 [Bacillus subtilis]	65	48	582
169	6	3192	3914	gi 1552737	similar to purine nucleoside phosphorylase (deoD) [Escherichia coli]	65	41	723
176	4	2931	2220	gnl PID e339500	oligopeptide binding lipoprotein [Streptococcus pneumoniae]	65	43	732
195	4	4556	3900	gi 1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	65	40	657
196	1	160	1572	gnl PID d102004	(AB001488) PROBABLE UDP-N-ACETYLURACINYL-D-GLUTAMYL-2, 6-DIAMINOLIGASE (EC 6.3.2.15). [Bacillus subtilis]	65	51	1413
204	2	2246	1215	gi 143156	membrane bound protein [Bacillus subtilis]	65	37	1032
210	4	1544	1891	gi 49315	ORF1 gene product [Bacillus subtilis]	65	48	348
242	2	1625	723	gi 1787540	(AE000226) f249; This 249 aa orf is 32 pct identical (8 gaps) to 244 residues of an approx. 272 aa protein AGAR_ECOLI SW: P42902 [Escherichia coli]	65	42	903

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
284	1	1	900	gi 559861	clYM [Plasmid pAD1]	65	36	900
304	1	2	574	gnl pid e290914	unknown [Mycobacterium tuberculosis]	65	52	573
315	1	2	1483	gi 790694	mannuronan C-5-epimerase [Azotobacter vinelandii]	65	57	1482
320	1	3	569	gnl pid d102048	K. aerogenes, histidine utilization repressor; p12180 (199) DNA binding [Bacillus subtilis]	65	46	567
358	1	1	309	gnl pid e323508	Y10S protein [Bacillus subtilis]	65	55	309
2	7	7571	6696	gi 1498753	nicotinate-nucleotide pyrophosphorylase [Rhodospirillum rubrum]	64	47	876
6	6	5924	6802	gnl pid d101111	methionine aminopeptidase [Synechocystis sp.]	64	52	879
8	4	3417	3686	gi 1045935	DNA helicase II [Mycoplasma genitalium]	64	58	270
11	4	3249	2689	gnl pid e265529	OrfB [Streptococcus pneumoniae]	64	46	561
15	7	6504	7145	gi 1762328	Ycr59c/Yig2 homolog [Bacillus subtilis]	64	45	642
22	11	9548	9895	gnl pid d100581	unknown [Bacillus subtilis]	64	38	348
22	30	22503	23174	gi 289260	comE ORF1 [Bacillus subtilis]	64	44	672
26	7	14375	14199	gi 409286	bmrU [Bacillus subtilis]	64	30	177
27	2	1310	1334	gi 40795	odei methylase [Desulfovibrio vulgaris]	64	51	177
29	2	614	297	gi 2326168	type VII collagen [Mus musculus]	64	50	318
35	2	368	721	pir JCL151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid Ti	64	50	354
40	1	3	449	gi 46970	epid gene product [Staphylococcus epidermidis]	64	41	447
40	7	4683	4976	gnl pid e325792	(AJ000005) glucose kinase [Bacillus megaterium]	64	45	294
45	7	8068	6920	gnl pid d102036	subunit of ADP-glucose pyrophosphorylase [Bacillus stearothermophilus]	64	40	1149
51	2	301	1059	gi 43985	infs-like gene [Lactobacillus delbrueckii]	64	54	759
51	13	15251	18397	gi 2293260	(AF008220) DNA-polymerase III alpha-chain [Bacillus subtilis]	64	46	3147
53	3	1157	555	gi 1574292	hypothetical [Haemophilus influenzae]	64	47	603
58	2	4236	1606	gi 1573826	alanine-tRNA synthetase (alaS) [Haemophilus influenzae]	64	51	2631
66	1	3	1259	gi 895749	putative cellobiose phosphotransferase enzyme II' [Bacillus subtilis]	64	42	1257
68	5	5213	6556	gi 436965	(maA) gene products [Bacillus stearothermophilus]	64	47	1344
69	6	5356	4949	gnl pid d101316	cdd [Bacillus subtilis]	64	52	408

TABLE 2
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
74	4	6948	5038	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]	64	50	1911
75	3	1283	1465	bbbs133379	TLS-CHOP-fusion protein(CHOP=C/EBP transcription factor, TLS=nuclear RNA-binding protein) [human, myxoid liposarcoma cells, Peptide Mutant, 462 aa] [Homo sapiens]	64	57	183
81	13	14016	14231	gi 143175	methanol dehydrogenase alpha-10 subunit [Bacillus sp.]	64	35	216
83	22	21851	22090	gnl pid d101315	YqfA [Bacillus subtilis]	64	44	240
87	11	10046	9300	gnl pid e323505	putative Ptd protein [Bacillus subtilis]	64	43	747
98	7	5032	5706	gnl pid e333880	hypothetical protein [Bacillus subtilis]	64	38	675
105	1	2	1276	gi 1657503	similar to S. aureus mercury(II) reductase [Escherichia coli]	64	45	1275
113	7	5136	6410	gnl pid d101119	Nifs [Synechocystis sp.]	64	50	1275
119	1	2	1297	gnl pid e320520	hypothetical protein [Natronobacterium pharaonis]	64	37	1296
123	3	1125	2156	gnl pid e353284	ORF YDL244w [Saccharomyces cerevisiae]	64	40	1032
124	5	2331	1780	gnl pid d101884	hypothetical protein [Synechocystis sp.]	64	50	552
129	4	3467	2709	gnl pid d101314	YqsU [Bacillus subtilis]	64	52	759
131	1	152	3	gi 1377841	unknown [Bacillus subtilis]	64	42	150
137	11	7196	7549	ptr JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS131) - Agrobacterium tumefaciens (strain P022) plasmid Ti	64	50	354
139	3	3226	2651	gi 2293301	(AF008220) YtqB [Bacillus subtilis]	64	44	576
146	10	6730	5648	gi 1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	64	45	1083
147	1	2	1018	gnl pid e137033	unknown gene product [Lactobacillus leichmannii]	64	46	1017
148	11	8430	8783	gi 2130630	(AF000430) dynamin-like protein [Homo sapiens]	64	28	354
156	7	4313	3612	gnl pid d102050	transmembrane [Bacillus subtilis]	64	31	702
157	4	1299	2114	gnl pid d100892	homologous to Gln transport system permease proteins [Bacillus subtilis]	64	43	816
162	6	5880	6362	gi 517204	ORF1, putative 42 kDa protein [Streptococcus pyogenes]	64	58	483
164	13	9707	8769	gnl pid d100964	homologue of ferric anguibactin transport system permease protein FatD of V. anguillarum [Bacillus subtilis]	64	40	939
175	5	3906	4598	gi 534045	antiterminator [Bacillus subtilis]	64	39	693
189	10	6154	6507	gi 581307	response regulator [Lactobacillus plantarum]	64	33	354
191	4	3519	2863	gi 149520	phosphoribosyl anthranilate isomerase [Lactococcus lactis]	64	46	657

TABLE 2
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
202	1	76	1140	[gnl]PTD e293806	O-acetylhomoserine sulphydrolase [Leptospira meyeri]	64	47	1065
224	1	234	1571	[gi]1573393	collagenase (prtC) [Haemophilus influenzae]	64	42	1338
231	3	291	647	[gi]40174	ORF X [Bacillus subtilis]	64	43	357
253	3	709	1089	[pir]J3C1151 J3C11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid Ti	64	50	381
265	1	820	2	[gi]1377832	unknown [Bacillus subtilis]	64	31	819
297	1	1	660	[gi]1590871	collagenase [Methanococcus jannaschii]	64	48	660
328	1	263	21	[gi]992651	Gln4p [Saccharomyces cerevisiae]	64	41	243
5	4	8730	8098	[gi]556885	unknown [Bacillus subtilis]	63	48	633
10	6	5178	4483	[gi]1573101	hypothetical [Haemophilus influenzae]	63	40	696
12	11	9324	9902	[gi]806536	membrane protein [Bacillus acidopullulyticus]	63	42	579
15	10	8897	9187	[gi]722339	unknown [Acetobacter xylinum]	63	40	291
17	2	1031	309	[gnl]PTD e217602	PinU [Lactobacillus plantarum]	63	32	723
18	8	7778	6975	[gi]1377843	unknown [Bacillus subtilis]	63	45	804
26	4	9780	7078	[gi]142440	ATP-dependent nuclease [Bacillus subtilis]	63	46	2703
29	5	3488	4192	[gi]1377829	unknown [Bacillus subtilis]	63	35	705
34	11	8830	7988	[gnl]PTD d101198	ORF8 [Enterococcus faecalis]	63	45	843
35	3	1187	876	[gi]722339	unknown [Acetobacter xylinum]	63	39	312
48	15	12509	11691	[gi]1573389	hypothetical [Haemophilus influenzae]	63	41	819
51	11	12719	12189	[gi]142450	ahrc protein [Bacillus subtilis]	63	35	531
55	4	3979	5022	[gi]1708640	YeaB [Bacillus subtilis]	63	41	1044
55	15	13669	14670	[gnl]PTD e11502	thioredoxine reductase [Bacillus subtilis]	63	44	1002
68	10	9242	8919	[sp]P37686 YIAY_	HYPOTHETICAL 40.2 KD PROTEIN IN AVTA-SELB INTERGENIC REGION (F382)	63	40	324
86	7	6554	5685	[gi]1574382	lic-1 operon protein (licD) [Haemophilus influenzae]	63	41	870
88	8	6085	5180	[gi]2098719	putative fimbrial-associated protein [Actinomyces naeslundii]	63	43	906
96	8	5858	6484	[gi]1052803	orf19yb gene product [Streptococcus pneumoniae]	63	38	627
100	1	240	1940	[gi]17171	glucosidase [Dictyostellium discoideum]	63	36	1701

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
104	4	3063	5765	gi1144985	phosphoenolpyruvate carboxylase [Corynebacterium glutamicum]	63	46	2703
106	8	9189	8554	gi1533099	endonuclease III [Bacillus subtilis]	63	45	636
122	6	4704	4886	gnl PID d101139	transposase [Synechocystis sp.]	63	39	183
128	7	4517	5203	gnl PID d101434	orf2 [Methanobacterium thermoautotrophicum]	63	50	687
137	4	963	1547	gi1472920	v-type Na-ATPase [Enterococcus hirae]	63	27	585
142	7	4100	4585	gnl PID e313025	hypothetical protein [Bacillus subtilis]	63	44	486
159	5	1741	2571	gi11787043	(AE000184) f271; This 271 aa orf is 24 pct identical (16 gaps) to 265 residues of an approx. 272 aa protein YIDA_ECOLI SW: P09997 [Escherichia coli]	63	39	831
171	12	8803	14406	gnl PID e224918	IgA1 protease [Streptococcus sanguis]	63	48	5604
177	1	3	347	gi11773150	hypothetical 14.8kd protein [Escherichia coli]	63	34	345
178	2	423	917	gi1722339	unknown [Acetobacter xylinum]	63	41	495
178	3	794	1012	gi11591582	cobalamin biosynthesis protein N [Methanococcus jannaschii]	63	36	219
195	1	1377	175	gnl PID e324217	ftsQ [Enterococcus hirae]	63	33	1203
234	5	1739	1527	gi11591582	cobalamin biosynthesis protein N [Methanococcus jannaschii]	63	36	213
249	1	81	257	gi11000453	TreR [Bacillus subtilis]	63	41	177
283	1	127	1347	gi1396486	ORF8 [Bacillus subtilis]	63	44	1221
293	3	2804	3466	gi1722339	unknown [Acetobacter xylinum]	63	37	663
311	1	905	486	gi11877424	UDP-galactose 4-epimerase [Streptococcus mutans]	63	46	420
324	1	2	556	gi11477741	histidine periplasmic binding protein P29 [Campylobacter jejuni]	63	36	555
365	1	219	13	gi12252843	(AF013293) No definition line found [Arabidopsis thaliana]	63	33	207
382	1	88	378	gi1722339	unknown [Acetobacter xylinum]	63	40	291
385	3	364	158	gi12252843	(AF013293) No definition line found [Arabidopsis thaliana]	63	33	207
2	1	2495	288	gnl PID e225007	penicillin-binding protein [Bacillus subtilis]	62	42	2208
3	23	23374	24231	gnl PID e254993	hypothetical protein [Bacillus subtilis]	62	35	858
6	16	14320	13193	gnl PID e349614	nifs-like protein [Mycobacterium leprae]	62	37	1128
7	8	6819	7232	gnl PID d101324	YqjY [Bacillus subtilis]	62	32	414
7	19	15466	14207	gnl PID d101804	beta ketoacyl-acyl carrier protein synthase [Synechocystis sp.]	62	43	1260

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
7	21	17155	16229	gnl PID e323514	putative FabD protein [Bacillus subtilis]	62	46	927
7	24	19526	18519	gi 1276434	beta-ketoacyl-ACP synthase III [Cuphea wrightii]	62	37	1008
12	7	5904	4702	gi 1573768	A/G-specific adenine glycosylase (mutV) [Haemophilus influenzae]	62	43	1203
12	9	8032	8793	gi 1591587	pantothenate metabolism flavoprotein [Methanococcus jannaschii]	62	33	762
15	11	9678	9328	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid Ti	62	43	351
17	4	2609	2442	gi 1591081	M. jannaschii predicted coding region MJ0374 [Methanococcus jannaschii]	62	43	168
17	5	3053	2835	gi 149570	role in the expression of lactacin F, part of the laf operon [Lactobacillus sp.]	62	44	219
22	10	8627	9538	gnl PID d100580	similar to B. subtilis DnaH [Bacillus subtilis]	62	43	912
30	3	865	2043	gi 2314379	[AE000627] ABC transporter, ATP-binding protein (yhG) [Helicobacter pylori]	62	43	1179
33	5	2235	1636	gi 413976	ipa-52r gene product [Bacillus subtilis]	62	44	600
38	11	5689	6123	gi 148231	o251 [Escherichia coli]	62	34	435
40	17	14272	13328	gnl PID d101904	hypothetical protein [Synechocystis sp.]	62	43	945
42	1	3	311	gi 1146182	putative [Bacillus subtilis]	62	41	309
44	2	1267	4005	gi 1786952	[AE000176] o877; 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment Y8GB_ECOLI SW: P54746 [Escherichia coli]	62	43	2739
48	12	9732	9304	gi 662920	repressor protein [Enterococcus hirae]	62	32	429
51	8	5664	7181	gnl PID e301153	StySKI methylase [Salmonella enterica]	62	44	1518
52	3	2791	2099	gi 1183886	integral membrane protein [Bacillus subtilis]	62	41	693
55	16	15702	14704	gnl PID e313028	hypothetical protein [Bacillus subtilis]	62	40	999
59	6	3418	3984	gi 2065483	unknown [Lactococcus lactis lactis]	62	32	567
63	5	4997	4809	gi 149771	pillin gene inverting protein (PivML) [Moraxella lacunata]	62	28	189
70	14	10002	10739	gi 992977	[bp]G gene product [Bordetella pertussis]	62	45	738
71	13	18790	20382	gi 1280135	coded for by C. elegans cDNA cm21e6; coded for by C. elegans cDNA cm01e2; similar to melibiose carrier protein (thiomethylgalactoside permease II) [Caenorhabditis elegans]	62	62	1593
71	28	32217	32768	gnl PID d101312	YqG [Bacillus subtilis]	62	35	552
74	7	11666	10383	gi 1552753	hypothetical [Escherichia coli]	62	38	1284

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
80	8	9370	9609	[gnl PID d102002]	(AB001488) FUNCTION UNKNOWN. [Bacillus subtilis]	62	46	240
97	10	9068	7041	[gi 882463]	protein-N(p1)-phosphohistidine-sugar phosphotransferase [Escherichia coli]	62	42	2028
98	4	2306	3268	[gnl PID d101496]	BrA8 (integral membrane protein) [Pseudomonas aeruginosa]	62	42	963
102	3	2823	3539	[gnl PID e13010]	hypothetical protein [Bacillus subtilis]	62	24	717
103	3	2795	1242	[gnl PID d102049]	H. influenzae hypothetical ABC transporter; P44808 (974) [Bacillus subtilis]	62	41	1554
111	2	2035	3462	[gi 581297]	MSP [Lactococcus lactis]	62	44	1428
112	4	3154	4080	[gi 1574379]	lic-1 operon protein (lica) [Haemophilus influenzae]	62	39	927
112	6	4939	5649	[gi 1574381]	lic-1 operon protein (licC) [Haemophilus influenzae]	62	39	711
124	3	1137	721	[gi 1573024]	anaerobic ribonucleoside-triphosphate reductase (nrtd) [Haemophilus influenzae]	62	45	417
124	6	3162	2329	[gi 609076]	leucyl aminopeptidase [Lactobacillus delbrueckii]	62	40	834
126	7	11073	7516	[gnl PID d101163]	ORF4 [Bacillus subtilis]	62	38	3558
129	6	4983	4540	[pir S41509 S415]	zinc finger protein EF6 - Chilo iridescent virus	62	48	444
131	7	4510	4103	[gi 1857245]	unknown [Lactococcus lactis]	62	42	408
149	2	1923	2579	[gi 1592142]	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	62	41	657
149	7	5360	6055	[gnl PID e323508]	YloS protein [Bacillus subtilis]	62	40	696
156	1	450	238	[gnl PID e254644]	membrane protein [Streptococcus pneumoniae]	62	40	213
156	6	3606	2935	[gnl PID d102050]	transmembrane [Bacillus subtilis]	62	37	672
171	2	1779	2291	[gi 43941]	ERII-B Sor PTS [Klebsiella pneumoniae]	62	35	513
172	2	385	723	[gi 895750]	putative cellobiose phosphotransferase enzyme III [Bacillus subtilis]	62	39	339
173	3	2599	893	[gi 1591732]	cobalt transport ATP-binding protein o [Methanococcus jannaschii]	62	42	1707
179	2	492	1754	[gi 1574071]	H. influenzae predicted coding region H1038 [Haemophilus influenzae]	62	38	1263
181	6	2856	3707	[gi 1777435]	Lact [Lactobacillus casei]	62	42	852
185	2	2074	311	[gi 2182397]	(AE000073) Y4fN [Rhizobium sp. NGR234]	62	41	1764
200	2	1061	1984	[gi 450566]	transmembrane protein [Bacillus subtilis]	62	37	924
202	3	2583	3473	[gi 42219]	p35 gene product (AA 1 - 314) [Escherichia coli]	62	41	891
210	3	1374	1565	[gi 49315]	ORF1 gene product [Bacillus subtilis]	62	45	192

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
211	1	3	971	gi1147402	mannose permease subunit III-Man (Escherichia coli)	62	43	969
223	2	1495	1034	gnl PID d101190	ORF2 [Streptococcus mutans]	62	41	462
228	1	34	909	gi1530063	glycerol uptake facilitator (Streptococcus pneumoniae)	62	44	876
234	2	90	917	gi12293259	(AF008220) YqJ [Bacillus subtilis]	62	38	828
282	5	1765	1487	gnl PID e276475	galactokinase (Arabidopsis thaliana)	62	33	279
375	1	1	159	gi11674231	(AE000052) Mycoplasma pneumoniae, hypothetical protein homolog; similar to Swiss-Prot Accession Number P35155, from B. subtilis (Mycoplasma pneumoniae)	62	40	159
385	5	584	357	gi11573353	outer membrane integrity protein (tolA) (Haemophilus influenzae)	62	47	228
3	19	18550	19269	gi1606162	ORF_f229 [Escherichia coli]	61	41	720
7	4	2725	3225	gi12114425	similar to Synchocystis sp. hypothetical protein, encoded by GenBank Accession Number D64006 [Bacillus subtilis]	61	42	501
17	6	3326	3054	gi1149589	lactacin F [Lactobacillus sp.]	61	43	273
44	3	4061	4957	gnl PID d101068	xylose repressor [Synchocystis sp.]	61	38	897
54	11	8388	7234	gnl PID d101329	YqjH [Bacillus subtilis]	61	42	1155
57	6	3974	6037	gnl PID d101316	YqfK [Bacillus subtilis]	61	42	2064
58	5	7356	6565	sp P45169 POTC	SPERMIDINE/PUTRESCINE TRANSPORT SYSTEM PERHEASE PROTEIN POTC	61	34	792
67	1	3	692	gi1537108	ORF_f254 [Escherichia coli]	61	46	690
68	9	8816	7890	gi119501	pPL212 gene product (AA 1-184) (Lupinus polyphyllus)	61	41	927
70	15	10737	12008	gi1392976	bplF gene product [Bordetella pertussis]	61	44	1272
72	11	9759	10202	gnl PID d101833	carboxynorspermidine decarboxylase [Synchocystis sp.]	61	36	444
76	8	7881	7003	gnl PID d100305	farnesyl diphosphate synthase [Bacillus stearothermophilus]	61	45	879
87	4	4914	3697	gi1528991	unknown [Bacillus subtilis]	61	42	1218
87	13	12311	11361	gi11789683	(AE000407) methionyl-tRNA formyltransferase [Escherichia coli]	61	44	951
91	2	731	2989	gi1537080	ribonucleoside triphosphate reductase [Escherichia coli]	61	45	2259
105	3	2711	3499	gnl PID d101851	hypothetical protein [Synchocystis sp.]	61	44	789
115	6	7968	6478	gi1895747	putative col operon regulator [Bacillus subtilis]	61	36	1491
123	8	7101	8518	gi1209527	protein histidine kinase (Enterococcus faecalis)	61	40	1338

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
126	6	7525	6725	gi 1787043	(AE000184) f271: This 271 aa orf is 24 pct identical (16 gaps) to 265 residues of an approx. 272 aa protein YIDA_ECOLI SW: P09997 [Escherichia coli]	61	38	801
128	1	1	639	gnl PID d101328	YgiY [Bacillus subtilis]	61	41	639
139	7	4794	5054	gi 1022726	unknown [Staphylococcus haemolyticus]	61	41	261
139	9	112632	5913	gnl PID e270014	beta-galactosidase [Thermotoga thermophilus]	61	41	6720
143	1	2552	42	gi 520541	penicillin-binding proteins 1A and 1B [Bacillus subtilis]	61	42	2511
148	16	12125	11424	gi 1552743	tetrahydrodipicolinate N-succinyltransferase [Escherichia coli]	61	42	702
162	3	4112	3456	gnl PID d101829	phosphoglycolate phosphatase [Synechocystis sp.]	61	30	657
172	3	727	1077	gnl PID d102048	B. subtilis, cellobiose phosphotransferase system, celsA; P46318 (220) [Bacillus subtilis]	61	44	351
177	3	1101	1772	gnl PID d100574	unknown [Bacillus subtilis]	61	43	672
202	2	1278	2585	gi 1045831	hypothetical protein (GB:L18965.6) [Mycoplasma genitalium]	61	36	1308
224	3	2782	3144	gi 1591144	M. jannaschii predicted coding region M30440 [Methanococcus jannaschii]	61	30	363
225	4	3395	3766	gi 1552774	hypothetical [Escherichia coli]	61	40	372
249	2	212	802	gi 1000453	Trer [Bacillus subtilis]	61	42	591
254	2	843	484	gnl PID d100417	ORF120 [Escherichia coli]	61	36	360
257	1	3	350	gnl PID e255315	unknown [Mycobacterium tuberculosis]	61	42	348
293	4	3971	3657	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid T1	61	45	315
301	1	949	17	gi 2291209	(AF016424) contains similarity to acyltransferases [Caenorhabditis elegans]	61	33	933
373	1	1066	287	gi 333396	Tb-292 membrane associated protein [Trypanosoma brucei subgroup]	61	38	780
3	24	24473	24955	gi 537093	ORF_0153b [Escherichia coli]	60	27	483
6	5	4636	5739	gi 2293258	(AF008220) YtoI [Bacillus subtilis]	60	35	1104
6	12	11936	11187	gi 293017	ORF3 (put.); putative [Lactococcus lactis]	60	44	750
17	13	6708	6484	gi 149569	lactacin F [Lactobacillus sp.]	60	32	225
18	7	6977	5670	gi 1788140	(AF000278) o481; This 481 aa orf is 35 pct identical (19 gaps) to 309 residues of an approx. 856 aa protein NOL1_HUMAN SW: P46087 [Escherichia coli]	60	43	1308
20	15	15878	17167	gnl PID d100584	unknown [Bacillus subtilis]	60	44	1290

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	1	1	243	[gnl PID d102050]	transmembrane [Bacillus subtilis]	60	36	243
32	10	8296	8964	[gi 2293275]	[AF008220] YtaG [Bacillus subtilis]	60	37	669
38	15	8837	9697	[gi 40023]	[B. subtilis] genes rnpH, rnpA, 50kd, gida and gida [Bacillus subtilis]	60	35	861
43	6	8610	5944	[gi 171787]	protein kinase 1 [Saccharomyces cerevisiae]	60	36	2667
44	1	1	1269	[gnl PID e235823]	unknown [Schizosaccharomyces pombe]	60	44	1269
45	10	11138	10368	[gi 397488]	1,4-alpha-glucan branching enzyme [Bacillus subtilis]	60	43	771
48	19	15766	14378	[gnl PID e205173]	orf1 [Lactobacillus helveticus]	60	39	1389
48	21	16727	16951	[gnl PID d102041]	unnamed protein product [Haemophilus actinomycetemcomitans]	60	32	225
50	1	2	898	[gnl PID e246537]	ORF286 protein [Pseudomonas stutzeri]	60	31	897
62	2	638	1177	[gnl PID d100587]	unknown [Bacillus subtilis]	60	42	540
68	4	3590	5203	[gi 1573583]	[H. influenzae] predicted coding region HI0594 [Haemophilus influenzae]	60	36	1614
70	11	5781	6182	[gnl PID d102014]	[AB001488] SIMILAR TO YDFR GENE PRODUCT OF THIS ENTRY (YDFR_BACSU).	60	33	402
70	12	6343	8133	[gnl PID e324970]	hypothetical protein [Bacillus subtilis]	60	38	1791
71	8	11701	14157	[gi 580866]	[ipa-12d] gene product [Bacillus subtilis]	60	33	2457
74	8	12509	11664	[gnl PID d101832]	phosphatidate cytidyltransferase [Synechocystis sp.]	60	45	846
76	4	4116	3367	[gi 2352096]	orf; similar to serine/threonine protein phosphatase [Pseudobacterium islandicum]	60	39	750
80	4	7372	7665	[gi 1786420]	[AE000131] f86; 100 pct identical to GB: ECODINJ_6 ACCESSION: D38582 [Escherichia coli]	60	30	294
81	6	4073	4522	[gi 147402]	mannose permease subunit III-Man [Escherichia coli]	60	35	450
86	1	940	155	[gi 143177]	putative [Bacillus subtilis]	60	26	786
92	1	1	192	[gi 396348]	homoserine transuccinylase [Escherichia coli]	60	45	192
93	14	10619	9384	[gi 1788389]	[AE000297] o464; This 464 aa orf is 33 pct identical (9 gaps) to 331 residues of an approx. 416 aa protein MTRC_NEIGO SW: P43505 [Escherichia coli]	60	27	1236
94	5	5548	8121	[gnl PID e329895]	cyclic nucleotide-gated channel beta subunit [Rattus norvegicus]	60	50	2574
97	7	5396	4523	[gi 1591396]	transketolase [Methanococcus jannaschii]	60	43	864
102	2	2081	2833	[gnl PID e320929]	hypothetical protein [Mycobacterium tuberculosis]	60	43	753

TABLE 2
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	9	9773	9183	gnl pid e334782	YibN protein [Bacillus subtilis]	60	31	591
113	8	6361	6837	gi 466675	nifU; B1496_C1_157 [Mycobacterium leprae]	60	43	477
115	2	2755	524	gnl pid e328143	(AJ000332) Glucosidase II (Homo sapiens)	60	32	2322
122	7	4763	5068	gnl pid d101876	transposase [Synecocystis sp.]	60	39	306
127	8	4510	5283	gi 1777938	Pgm (Treponema pallidum)	60	38	774
138	4	3082	2672	gnl pid e325196	hypothetical protein [Bacillus subtilis]	60	36	411
139	1	177	4	gnl pid d100680	ORF [Thermus thermophilus]	60	39	174
139	11	14520	13009	gi 537145	ORF_f437 [Escherichia coli]	60	30	1512
140	2	2592	1249	gi 1209527	protein histidine kinase [Enterococcus faecalis]	60	37	1344
141	1	210	1049	gi 463181	ES ORF from bp 3842 to 4081; putative [Human papillomavirus type 33]	60	34	840
141	5	5368	6405	gi 145362	tyrosine-sensitive DAP synthase (aroF) [Escherichia coli]	60	41	1038
142	6	3558	4049	gi 600711	putative [Bacillus subtilis]	60	37	492
148	10	7742	8713	gnl pid e313022	hypothetical protein [Bacillus subtilis]	60	27	972
153	5	3667	4278	gi 2293332	(AF008220) branch-chain amino acid transporter [Bacillus subtilis]	60	42	612
155	1	1413	748	gi 2104504	putative UDP-glucose dehydrogenase [Escherichia coli]	60	40	666
158	3	3116	2472	gnl pid d100872	a negative regulator of pho regulon [Pseudomonas aeruginosa]	60	37	645
159	3	778	1386	gnl pid e308090	product highly similar to Bacillus anthracis CapA protein [Bacillus subtilis]	60	48	609
163	7	8049	8468	gnl pid d101313	Yqen [Bacillus subtilis]	60	38	420
170	3	4130	2688	gi 1574179	H. influenzae predicted coding region HI1244 [Haemophilus influenzae]	60	39	1443
171	7	4717	5901	gi 606076	ORF_0384 [Escherichia coli]	60	44	1185
183	3	2440	2135	gi 1877427	repressor [Streptococcus pyogenes phage T12]	60	38	306
191	10	9444	8428	gi 415664	catabolite control protein [Bacillus megaterium]	60	42	1017
200	1	139	1083	gi 438462	transmembrane protein [Bacillus subtilis]	60	37	945
201	3	3695	1928	gi 475112	enzyme Iabc (Pedococcus pentosaceus)	60	39	1968
214	15	10930	10439	gi 1573407	hypothetical [Haemophilus influenzae]	60	39	492
218	4	2145	2363	gi 608520	myosin heavy chain kinase A [Dictyostelium discoideum]	60	31	219

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
226	4	2518	2351	gi 437705	hyaluronidase [Streptococcus pneumoniae]	60	53	168
242	1	725	3	gi 43938	Sor regulator [Klebsiella pneumoniae]	60	41	723
245	1	1	288	gi 304897	EcoE type I restriction modification enzyme H subunit [Escherichia coli]	60	56	288
251	1	905	45	gi 671632	unknown [Staphylococcus aureus]	60	36	861
259	1	969	82	gi 153794	rgg [Streptococcus gordonii]	60	32	888
260	2	1492	1662	gi 31840 S318	probable transposase - Bacillus stearothermophilus	60	26	171
274	1	836	96	gi 1592173	N-ethylmaleimide chlorohydrolase [Methanococcus jannaschii]	60	40	741
308	1	463	2	gi 1787397	(AE00214) o157 [Escherichia coli]	60	43	462
318	1	3	308	gnl pid e137594	xerC recombinase [Lactobacillus leichmannii]	60	42	306
344	1	73	522	gi 509672	repressor protein [Bacteriophage Tuc2009]	60	32	450
5	1	576	4	gi 2293147	(AF008220) YtxM [Bacillus subtilis]	59	31	573
7	22	18140	17142	gnl pid e280724	unknown [Mycobacterium tuberculosis]	59	39	999
10	1	1413	4	gi 1353880	sialidase L [Macrobodella decora]	59	41	1410
15	6	6463	5156	gi 580841	F1 [Bacillus subtilis]	59	35	1308
22	2	479	1393	gi 142469	ale operon regulatory protein [Bacillus subtilis]	59	34	915
22	5	2698	4614	gnl pid e280623	PCPA [Streptococcus pneumoniae]	59	44	1917
30	1	208	558	gnl pid e233868	hypothetical protein [Bacillus subtilis]	59	37	351
30	4	3678	2455	gnl pid e202290	unknown [Lactobacillus sake]	59	33	1224
35	13	12201	11071	gnl pid e238664	hypothetical protein [Bacillus subtilis]	59	35	1131
35	14	13288	12182	gi 1657647	Cap8M [Staphylococcus aureus]	59	39	1107
36	18	18076	17897	gi 1500535	M. jannaschii predicted coding region MJ1635 [Methanococcus jannaschii]	59	33	180
38	12	6172	7137	gi 2293239	(AF008220) YtxK [Bacillus subtilis]	59	34	966
42	3	1952	3361	gi 1684845	pilin [Canis familiaris]	59	40	1410
50	3	2678	1728	gnl pid d101329	VqjK [Bacillus subtilis]	59	41	951
56	5	1870	2388	gnl pid e137594	xerC recombinase [Lactobacillus leichmannii]	59	41	519
61	6	6812	5628	gnl pid e311516	aminotransferase [Bacillus subtilis]	59	40	1185
67	5	2382	3023	gi 1146190	2-keto-3-deoxy-6-phosphogluconate aldolase [Bacillus subtilis]	59	36	642

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
69	10	8567	8899	gi 1573628	antiothenate kinase (coaA) [Haemophilus influenzae]	59	38	333
87	12	11383	10055	gnl PID e323504	putative Fnu protein [Bacillus subtilis]	59	44	1329
113	14	13927	15894	gi 1673731	(AE000010) Mycoplasma pneumoniae, fructose-permease IIBC component; similar to Swiss-Prot Accession Number P20966, from E. coli [Mycoplasma pneumoniae]	59	43	1968
115	8	8766	8521	gi 1590886	M. jannaschii predicted coding region MJ0110 [Methanococcus jannaschii]	59	38	246
119	2	1966	1526	gnl PID e209005	homologous to ORF2 in nrDEF operons of E. coli and S. typhimurium [Lactococcus lactis]	59	43	441
128	17	13438	13178	gnl PID e279632	unknown [Mycobacterium tuberculosis]	59	38	261
140	22	23903	23388	gi 482922	protein with homology to pail repressor of B. subtilis [Lactobacillus delbrueckii]	59	40	516
148	13	9697	9014	gnl PID d102005	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN H. INFLUENZAE AND SYNECHOCYSTIS. [Bacillus subtilis]	59	32	684
149	10	7213	8244	gi 710422	cmp-binding-factor 1 [Staphylococcus aureus]	59	40	1032
164	9	6993	6013	gnl PID d100965	ferric anguibactin-binding protein precursor PatB of V. anguillarum [Bacillus subtilis]	59	41	981
164	12	8836	7823	gnl PID d100964	homologue of ferric anguibactin transport system permease protein PatC of V. anguillarum [Bacillus subtilis]	59	35	1014
177	2	401	1072	gi 289759	coded for by C. elegans CDNA CE2G3 (GenBank:Z14728); putative [Caenorhabditis elegans]	59	40	672
177	7	3841	4200	gi 2313445	(AE000551) H. pylori predicted coding region HP0342 [Helicobacter pylori]	59	38	360
183	4	2768	2508	gi 509672	repressor protein [Bacteriophage Tuc2009]	59	50	261
186	6	3398	2820	gi 606080	ORF_0290; Genepilot suggests frameshift linking to o267, not found [Escherichia coli]	59	38	579
190	3	3120	1711	gi 1613768	histidine protein kinase [Streptococcus pneumoniae]	59	32	1410
194	2	1621	1019	gnl PID d100579	unknown [Bacillus subtilis]	59	40	603
198	7	5205	4306	gnl PID e313073	hypothetical protein [Bacillus subtilis]	59	38	900
220	5	4362	3958	gnl PID d101322	YqjL [Bacillus subtilis]	59	46	405
242	3	1573	2367	gi 1787045	(AE000184) f308; this 308 aa orf is 35 pct identical (35 gaps) to 305 residues of an approx. 296 aa protein PFLC_ECOLI SW: P32675 [Escherichia coli]	59	42	795
247	2	1154	1480	gi 40073	ORF107 [Bacillus subtilis]	59	39	327

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
256	1	868	2	gml PID d101924	hemolysin [Synechocystis sp.]	59	39	867
258	1	65	820	gi 2246532	ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]	59	20	756
270	1	386	1126	gml PID d102092	YfnB [Bacillus subtilis]	59	40	741
281	1	552	166	gi 666062	putative [Lactococcus lactis]	59	31	387
309	1	3	479	gi 405879	yleH [Escherichia coli]	59	38	477
363	1	2	1894	gi 915208	gastric mucin [Sus scrofa]	59	31	1893
387	2	425	84	gi 160671	S antigen precursor [Plasmodium falciparum]	59	44	342
5	6	11223	10465	gml PID d101812	LumQ [Synechocystis sp.]	58	29	759
29	4	2098	3513	gml PID d100479	Na ⁺ -ATPase subunit J [Enterococcus hirae]	58	39	1416
30	5	4058	3651	gi 39478	ATP binding protein of transport ATPases [Bacillus firmus]	58	34	408
33	6	2983	2210	gml PID d101164	unknown [Bacillus subtilis]	58	45	774
36	8	5316	6179	gi 1518679	orf [Bacillus subtilis]	58	32	864
43	5	5926	3971	gi 1788150	[AE000278] protease II [Escherichia coli]	58	37	1956
46	5	3704	5221	gml PID e267329	Unknown [Bacillus subtilis]	58	42	1518
48	14	11722	11066	gml PID d101771	thiamin biosynthetic bifunctional enzyme [Synechocystis sp.]	58	34	657
52	1	1229	3	gml PID d101291	reductase [Pseudomonas aeruginosa]	58	35	1227
53	2	702	412	gi 2313357	[AE000545] cytochrome c biogenesis protein (ccda) [Helicobacter pylori]	58	25	291
58	4	6586	5498	gi 147329	transport protein [Escherichia coli]	58	41	1089
69	5	4934	3807	gml PID e311492	unknown [Bacillus subtilis]	58	41	1128
71	27	31357	32277	gi 2408014	hypothetical protein [Schizosaccharomyces pombe]	58	33	921
72	4	3586	2882	gi 18694	nodulin-21 (AA 1-201) [Glycine max]	58	34	705
74	3	4937	4230	gi 2293252	[AF008220] Ymo [Bacillus subtilis]	58	33	708
79	4	4594	3422	gi 1217989	[ORF3] [Streptococcus pneumoniae]	58	44	1173
82	8	10585	8171	gi 882711	exonuclease V alpha-subunit [Escherichia coli]	58	38	2415
86	17	16017	15337	gi 47642	5-dehydroquinase hydrolyase (3-dehydroquinase) [Salmonella typhi]	58	32	681
97	2	931	560	gi 153794	rgg [Streptococcus gordonii]	58	32	372

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
108	2	358	2724	gi 537020	vacB gene product (Escherichia coli)	58	37	2367
111	5	4593	5240	gi 1592142	ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)	58	36	648
120	3	4421	5110	gnl PID d101320	YggX (Bacillus subtilis)	58	47	690
128	16	13131	12673	gi 662919	ORF U (Enterococcus hirae)	58	42	459
132	3	6174	4939	gi 1800301	macrolide-efflux determinant (Streptococcus pneumoniae)	58	35	1236
133	1	111	890	gnl PID e269488	Unknown (Bacillus subtilis)	58	36	780
160	11	8615	9865	gi 473901	ORF1 (Lactococcus lactis)	58	39	1251
161	6	6268	6849	gnl PID d101024	DJ-1 protein (Homo sapiens)	58	32	582
169	1	214	2	gnl PID d100447	translation elongation factor-3 (Chlorella virus)	58	31	213
187	1	487	2	gi 475114	regulatory protein (Pediococcus pentosaceus)	58	38	486
187	6	4384	4620	gi 167475	desiccation-related protein (Craterostigma plantagineum)	58	55	237
190	2	1464	1640	gnl PID e246727	competence pheromone (Streptococcus gordonii)	58	38	177
192	2	2012	1344	gnl PID d100556	rat GCP360 (Rattus rattus)	58	44	669
206	1	1292	696	gnl PID e205579	product similar to WrbA (Lactobacillus sake)	58	35	597
216	2	2333	555	gnl PID e325036	hypothetical protein (Bacillus subtilis)	58	33	1779
217	5	5250	4321	gi 466474	cellobiose phosphotransferase enzyme II'' (Bacillus stearothermophilus)	58	38	930
217	7	5636	5106	gnl PID d102048	B. subtilis cellobiose phosphotransferase system celB; P46317 (998) transmembrane (Bacillus subtilis)	58	44	531
232	1	2	811	gi 1573777	cell division ATP-binding protein (ftsE) (Haemophilus influenzae)	58	39	810
264	1	2	715	gi 973330	NatA (Bacillus subtilis)	58	32	714
280	1	33	767	gi 1786187	(AE000111) hypothetical 29.6 kD protein in thrC-talB intergenic region (Escherichia coli)	58	31	735
306	1	845	3	gnl PID e334780	YibL protein (Bacillus subtilis)	58	47	843
360	3	1556	1092	sp P46351 YEGD_	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'-REGION.	58	32	465
363	5	2160	1867	gi 160671	S antigen precursor (Plasmodium falciparum)	58	51	294
372	1	806	3	gi 393394	Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	58	37	804
382	2	749	519	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS131) - Agrobacterium tumefaciens (strain P022) plasmid T1	58	41	231

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3	9	8409	7471	gi 1499745	M. jannaschii predicted coding region MJ0912 [Methanococcus jannaschii]	57	38	939
10	10	7674	7507	gi 1737169	homologue to SKP1 [Arabidopsis thaliana]	57	30	168
11	1	2	412	gnl PID d100139	ORF [Acetobacter pasteurianus]	57	42	411
31	4	2032	1388	gi 2293213	[AF008220] YtpR [Bacillus subtilis]	57	37	645
33	11	6931	6449	gnl PID d124949	hypothetical protein [Bacillus subtilis]	57	36	483
45	5	5446	5060	gi 1592204	phosphoserine phosphatase [Methanococcus jannaschii]	57	44	387
49	7	6523	7632	gi 155369	PTS enzyme-II fructose [Xanthomonas campestris]	57	35	1110
52	6	4520	6850	gi 1574144	single-stranded-DNA-specific exonuclease (recJ) [Haemophilus influenzae]	57	35	2331
53	5	2079	1795	gi 1843580	replicase-associated polypeptide (coat blue dwarf virus)	57	46	285
63	6	5312	4995	gi 2182608	[AE000094] Y4rJ [Rhizobium sp. NGR234]	57	39	318
72	15	13883	13059	gnl PID d100892	homologous to SwissProt:YIDA_ECOLI hypothetical protein [Bacillus subtilis]	57	40	825
79	2	2561	1815	gnl PID d100965	homologue of NADPH-flavin oxidoreductase Fp of V. harveyi [Bacillus subtilis]	57	44	747
82	9	9596	9763	gi 1206045	short region of similarity to glycerophosphoryl diester phosphodiesterases [Caenorhabditis elegans]	57	35	168
86	16	15371	14493	gi 1787983	[AE000264] o288; 92 pct identical (1 gap) to 222 residues of fragment YDIB_ECOLI SW: F28244 (223 aa) [Escherichia coli]	57	34	879
93	3	1695	1177	gi 1500003	mutator mutT protein [Methanococcus jannaschii]	57	33	519
96	6	3026	4519	gi 1559882	threonine synthase [Arabidopsis thaliana]	57	43	1494
99	14	17211	18212	gi 773349	BirA protein [Bacillus subtilis]	57	44	1002
112	8	7448	7903	gi 1591393	M. jannaschii predicted coding region MJ0678 [Methanococcus jannaschii]	57	30	456
113	16	18627	18328	pir A45605/A456	mature-parasite-infected erythrocyte surface antigen WESA - Plasmodium falciparum	57	22	300
123	2	343	1110	pir F64149/F641	hypothetical protein HI0355 - Haemophilus influenzae (strain Rd KW20)	57	38	768
123	4	2108	2884	gnl PID d102148	[AB001684] sulfate transport system permease protein [Chlorella vulgaris]	57	39	777
127	10	6477	5587	gi 1573082	nitrogenase C (nifC) [Haemophilus influenzae]	57	35	891
128	13	9251	9790	gi 153692	pneumolysin [Streptococcus pneumoniae]	57	38	540
131	4	2139	1363	gi 42081	magD gene product (AA 1-250) [Escherichia coli]	57	36	777

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
136	1	214	1221	bbs148453	Spaa-endocarditis immunodominant antigen (Streptococcus sobrinus), MUCOB 263, Peptide, 1566 aa [Streptococcus sobrinus]	57	44	1008
140	25	28701	26851	gi1505576	beta-glucoside permease [Bacillus subtilis]	57	38	1851
141	6	6395	7438	gi1995560	unknown [Schizosaccharomyces pombe]	57	41	1044
144	3	3231	2785	gnlpid100139	ORF [Acetobacter pasteurianus]	57	42	447
155	4	5454	4564	gi1600431	glycosyl transferase [Erwinia amylovora]	57	34	891
159	9	4877	5854	gi1290509	o307 [Escherichia coli]	57	35	978
167	11	9710	9249	gnlpid100139	ORF [Acetobacter pasteurianus]	57	42	462
171	6	4023	4436	gi147402	mannose permease subunit III-Man [Escherichia coli]	57	29	414
178	4	2170	1076	gnlpid102004	(AB001488) ATP-DEPENDENT RNA HELICASE DEAD HOMOLOG. [Bacillus subtilis]	57	39	1095
190	1	145	1455	gi149420	export/processing protein [Lactococcus lactis]	57	30	1311
198	1	298	95	gi1522268	unidentified ORF22 [Bacteriophage BIL67]	57	36	204
203	2	3195	2110	gnlpid10283915	orf c01003 [Sulfolobus solfataricus]	57	41	1086
205	1	40	507	gi1439527	ETIA-man [Lactobacillus curvatus]	57	28	468
214	7	4243	3797	gnlpid102049	H. influenzae, ribosomal protein alanine acetyltransferase; P44305 (189) [Bacillus subtilis]	57	48	447
268	3	1767	1276	gi143979	L. curvatus small cryptic plasmid gene for rep protein [Lactobacillus curvatus]	57	36	492
351	1	324	34	gnlpid10275871	T03F6.b [Caenorhabditis elegans]	57	31	291
386	1	226	2	gi1160671	S antigen precursor [Plasmodium falciparum]	57	45	225
5	5	10486	8777	gi1405857	yehu [Escherichia coli]	56	33	1710
8	5	3674	3910	gi1467199	pksC; L518.F1.2 [Mycobacterium leprae]	56	39	237
10	3	3442	1874	gnlpid101907	sodium-coupled permease [Synecocystis sp.]	56	36	1569
21	1	1880	333	gi12313949	(AE000593) osmoprotection protein (proWX) [Helicobacter pylori]	56	33	1548
22	29	21968	22456	gnlpid102001	(AB001488) PROBABLE ACETYLTRANSFERASE. [Bacillus subtilis]	56	37	489
27	1	1361	3	gi1215132	ea59 (525) [Bacteriophage lambda]	56	30	1359
28	9	4667	4278	gi1592090	DNA repair protein RAD2 [Methanococcus jannaschii]	56	29	390
33	1	3	386	gnlpid100139	ORF [Acetobacter pasteurianus]	56	41	384

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
36	7	5122	5397	pir PQ0053 PQ00	hypothetical protein (proc 3' region) - <i>Pseudomonas aeruginosa</i> (strain PAO) (fragment)	56	28	276
40	4	3137	4318	gi 1800301	macrolide-efflux determinant [<i>Streptococcus pneumoniae</i>]	56	27	1182
40	16	12511	13191	gnl PID e217602	PinU [<i>Lactobacillus plantarum</i>]	56	38	681
48	17	13775	13023	gi 143729	transcription activator [<i>Bacillus subtilis</i>]	56	35	753
75	4	1674	2594	gnl PID d102036	membrane protein [<i>Bacillus stearothermophilus</i>]	56	25	921
85	3	1842	1459	gnl PID d100139	ORF [<i>Acetobacter pasteurianus</i>]	56	41	384
89	7	5815	4940	gi 853777	product similar to <i>E. coli</i> PRP2 protein [<i>Bacillus subtilis</i>]	56	42	876
105	2	1160	2718	gnl PID d101913	hypothetical protein [<i>Synechocystis</i> sp.]	56	37	1359
112	3	2151	3194	gi 537201	ORF_0345 [<i>Escherichia coli</i>]	56	31	1044
113	4	2754	2963	gnl PID d100340	ORF [Plum pox virus]	56	28	210
122	3	1203	2054	gi 1649035	high-affinity periplasmic glutamine binding protein [<i>Salmonella typhimurium</i>]	56	30	852
124	8	3939	3694	gnl PID e248893	unknown [<i>Mycobacterium tuberculosis</i>]	56	27	246
125	4	4403	4107	gnl PID d100247	human non-muscle myosin heavy chain (Homo sapiens)	56	32	297
127	11	6608	6405	gi 2182397	(AE000073) Y4fn [<i>Rhizobium</i> sp. NGR234]	56	35	204
134	5	4769	3849	gnl PID d101870	hypothetical protein [<i>Synechocystis</i> sp.]	56	39	921
137	10	6814	7245	gi 1592011	sulfate permease (cysA) [<i>Methanococcus jannaschii</i>]	56	34	432
142	8	5019	4582	pir A47071 A470	orf1 immediately 5' of nifs - <i>Bacillus subtilis</i>	56	29	438
146	8	4676	3660	gnl PID d101911	hypothetical protein [<i>Synechocystis</i> sp.]	56	32	1017
148	3	1906	2739	gnl PID d101099	phosphate transport system permease protein PstA [<i>Synechocystis</i> sp.]	56	36	834
150	4	4449	2743	gnl PID e304628	probably site-specific recombinase of the resolvase family of enzymes [<i>Bacteriophage TP21</i>]	56	27	1707
172	1	2	208	gi 1787791	(AE000249) f317; This 317 aa orf is 27 pct identical (16 gaps) to 301 residues of an approx. 320 aa protein YKKC_BACSU SW: F39140 [<i>Escherichia coli</i>]	56	34	207
172	7	4979	5668	gi 396293	similar to <i>Bacillus subtilis</i> hypoth. 20 kDa protein, in tsr 3' region [<i>Escherichia coli</i>]	56	40	690
186	7	3732	3367	gi 1732200	PTS permease for mannose subunit IIPMan [<i>Vibrio furnissii</i>]	56	36	366
187	2	2402	819	pir S57904 S579	vir49 protein - <i>Streptococcus pyogenes</i> (strain CS101, serotype M49)	56	35	1584

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
204	3	2772	2239	gi 606376	ORF_ol62 [Escherichia coli]	56	35	534
206	2	3342	1633	gi 559861	plyM [Plasmid pAD1]	56	38	1710
219	3	1689	1096	gi 1146197	putative [Bacillus subtilis]	56	27	594
230	2	409	1485	pir C60328 C603	hypothetical protein 2 (sr 5' region) - Streptococcus mutans (strain OH2175, serotype f)	56	40	1077
233	4	2930	3268	gi 1041785	rhostry protein [Plasmodium yoelii]	56	24	339
273	2	1543	2724	gi 143089	lep protein [Bacillus subtilis]	56	32	1182
353	1	1	516	gnl PID e325000	hypothetical protein [Bacillus subtilis]	56	41	516
359	1	87	641	gi 1786952	(AE000176) o877, 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment Y8GB_ECOLI SW: P54746 [Escherichia coli]	56	46	555
363	7	4482	4198	gi 1573353	outer membrane integrity protein (tolA) [Haemophilus influenzae]	56	38	285
376	1	2	508	gnl PID e325031	hypothetical protein [Bacillus subtilis]	56	33	507
18	1	836	177	gnl PID d100872	a negative regulator of pho regulon [Pseudomonas aeruginosa]	55	31	660
28	4	1824	1618	gnl PID e316518	STAR protein [Dictyostelium discoideum]	55	40	207
29	6	4496	5041	gi 1088261	unknown protein [Anabaena sp.]	55	31	546
38	16	9695	10702	gi 580905	B. subtilis genes rpmH, rpmA, 50kd, gidA and gidB [Bacillus subtilis]	55	31	1008
49	5	5727	6182	gi 1786951	(AE000176) heat-responsive regulatory protein [Escherichia coli]	55	29	456
51	4	2381	3241	gnl PID d101293	ybbA [Bacillus subtilis]	55	42	861
52	9	9640	10866	gi 153016	ORF_419 protein [Staphylococcus aureus]	55	23	1227
53	4	1813	1349	gi 896042	OspF [Borrelia burgdorferi]	55	30	465
60	5	4794	5756	gi 1499876	magnesium and cobalt transport protein [Methanococcus jannaschii]	55	38	963
71	9	14176	15408	gi 1857120	glycosyl transferase [Neisseria meningitidis]	55	41	1233
75	6	3189	4229	gnl PID e209890	NAD alcohol dehydrogenase [Bacillus subtilis]	55	44	1041
108	10	10488	9820	gnl PID e324997	hypothetical protein [Bacillus subtilis]	55	36	669
113	12	12273	13037	gnl PID e311496	unknown [Bacillus subtilis]	55	34	765
113	13	13007	13945	gi 1573423	1-phosphofructokinase (fruk) [Haemophilus influenzae]	55	39	939
126	5	6764	5907	gi 1790131	(AE000446) hypothetical 29.7 kD protein in ibpA-gyrB intergenic region [Escherichia coli]	55	37	858

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
129	3	2719	902	gnl PID d101425	Pz-peptidase (Bacillus licheniformis)	55	35	1818
138	3	2593	1610	gi 142833	ORF2 (Bacillus subtilis)	55	37	984
140	6	6916	5633	gnl PID d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus (Bacillus subtilis)	55	26	1284
147	3	3854	2136	gi 472330	dlhydroxypamide dehydrogenase (Clostridium magnum)	55	39	1719
147	10	10204	8921	gnl PID e73078	dlhydroxypamide dehydroxypamide dehydroxypamide (Lactobacillus leichnamnii)	55	38	1284
148	5	3430	4119	gi 290572	peripheral membrane protein U (Escherichia coli)	55	29	690
148	6	4171	4650	gi 695769	transposase (Xanthobacter autotrophicus)	55	37	480
149	14	12564	11650	gnl PID d101329	YqjG (Bacillus subtilis)	55	32	915
156	3	1113	550	gi 2314496	(AE000634) conserved hypothetical integral membrane protein (Helicobacter pylori)	55	34	564
159	10	6625	5897	gi 290533	similar to E. coli ORF adjacent to suc operon; similar to gntR class of regulatory proteins (Escherichia coli)	55	29	729
164	3	1784	2332	gnl PID e55118	hypothetical protein (Bacillus subtilis)	55	37	549
164	5	2772	3521	gi 40348	put. resolvase Tnp I (AA 1 - 284) (Bacillus thuringiensis)	55	35	750
164	11	7428	7216	gnl PID e249407	unknown (Mycobacterium tuberculosis)	55	38	213
167	5	3860	3345	gi 535052	involved in protein secretion (Bacillus subtilis)	55	28	516
186	5	2880	2563	gi 606080	ORF_0290; Geneplot suggests frameshift linking to o267, not found (Escherichia coli)	55	35	318
189	8	4311	5396	gnl PID e183450	hypothetical EcsB protein (Bacillus subtilis)	55	32	1086
192	5	3270	3079	gi 1196504	vitellogenin convertase (Aedes aegypti)	55	38	192
195	2	2454	1384	gi 1574693	transferase, peptidoglycan synthesis (murG) (Haemophilus influenzae)	55	33	1071
198	4	3013	2471	gnl PID e313074	hypothetical protein (Bacillus subtilis)	55	29	543
214	1	373	744	gnl PID d101741	transposase (Synecocystis sp.)	55	33	372
219	2	1115	456	gi 288301	ORF2 gene product (Bacillus megaterium)	55	30	660
263	7	3742	3443	gi 18137	cgcr-4 product (Chlamydomonas reinhardtii)	55	48	300
285	1	2	829	gnl PID d100974	unknown (Bacillus subtilis)	55	40	828
286	1	650	249	gi 396844	ORF (18 kDa) (Vibrio cholerae)	55	31	402
297	2	1229	1696	gi 150848	prtC (Porphyromonas gingivalis)	55	39	468

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
309	2	218	982	gi 1574491	hypothetical [Haemophilus influenzae]	55	35	765
328	2	646	224	gi 571500	prohibitin [Saccharomyces cerevisiae]	55	27	423
330	1	1340	474	gi 396397	soxS [Escherichia coli]	55	29	867
364	3	2338	1546	gi 393394	Tb-291 membrane associated protein [Trypanosoma brucei subgroup]	55	36	993
368	3	941	105	gi 160671	S antigen precursor [Plasmodium falciparum]	55	40	837
3	5	4604	3624	gi 2293176	[AF008220] signal transduction protein kinase [Bacillus subtilis]	54	26	981
9	11	7746	7246	gi 1146245	putative [Bacillus subtilis]	54	38	501
38	24	16213	17937	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	54	27	1725
40	8	5076	4882	gi 39989	methionyl-tRNA synthetase [Bacillus stearothermophilus]	54	35	195
43	4	3980	2367	gnl PID e148611	ABC transporter [Lactobacillus helveticus]	54	25	1614
52	10	10844	12103	gi 1762962	FenA [Staphylococcus simulans]	54	29	1260
57	1	3	512	gi 558177	endo-1,4-beta-xylanase [Cellulomonas fimi]	54	36	510
58	3	4749	4246	gnl PID d101237	hypothetical [Bacillus subtilis]	54	29	504
71	7	10684	11703	gi 510255	orf3 [Escherichia coli]	54	31	1020
71	20	27546	27737	gi 202543	serotonin receptor [Rattus norvegicus]	54	31	192
72	2	844	1098	gi 148613	srnB gene product [Plasmid F]	54	37	255
72	7	7438	6695	gi 1196496	recombinase [Moraxella bovis]	54	38	744
74	10	14043	13465	gi 1200342	ORF 3 gene product [Bradyrhizobium japonicum]	54	32	579
74	12	16483	15995	gi 2317798	maturase-related protein [Pseudomonas alcaligenes]	54	30	489
86	3	2877	2155	gi 46988	orf9.6 possibly encodes the O unit polymerase [Salmonella enterica]	54	34	723
89	5	4433	3921	gi 147211	phnO protein [Escherichia coli]	54	41	513
90	1	3	464	gi 2317798	maturase-related protein [Pseudomonas alcaligenes]	54	30	462
96	10	8058	8510	gnl PID d102015	[AB001488] SIMILAR TO SALMONELLA TYPHIMURUM SLVY GENE REQUIRED FOR SURVIVAL IN MACROPHAGE. [Bacillus subtilis]	54	32	453
97	6	4662	3604	gi 1591394	transketolase [Methanococcus jannaschii]	54	30	1059
106	11	10406	12010	gi 606286	ORF_0637 [Escherichia coli]	54	32	1605
147	8	8663	7404	gnl PID d101615	ORF_ID:031987; similar to [SwissProt Accession Number P37340] [Escherichia coli]	54	35	1260

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
171	4	2477	3223	gi 1439528	EIIC-man [Lactobacillus curvatus]	54	36	747
174	2	2068	1787	gnl PID d100518	motor protein [Homo sapiens]	54	35	282
188	1	526	1188	gnl PID e250352	unknown [Mycobacterium tuberculosis]	54	31	663
198	5	3582	2884	gnl PID e313074	hypothetical protein [Bacillus subtilis]	54	33	699
207	1	1	1641	gnl PID d101813	hypothetical protein [Synecocystis sp.]	54	24	1641
210	1	2	655	gi 2293206	[AF008220] Ymp [Bacillus subtilis]	54	29	654
225	2	966	2357	gnl PID e330194	rl1186.1 [Caenorhabditis elegans]	54	39	1392
241	1	1681	347	gnl PID d101813	hypothetical protein [Synecocystis sp.]	54	26	1335
263	2	907	1395	gnl PID d101886	transposase [Synecocystis sp.]	54	30	489
263	6	3450	2977	gi 160671	S antigen precursor [Plasmodium falciparum]	54	47	474
277	3	2517	1363	gi 1196926	unknown protein [Streptococcus mutans]	54	30	1155
307	1	828	4	gi 2293198	[AF008220] Ytgp [Bacillus subtilis]	54	28	825
325	1	19	768	gi 2182507	[AE000083] Y41H [Rhizobium sp. NGR234]	54	37	750
332	2	898	590	gi 1591815	ADP-ribosylglycohydrolase [drao] [Methanococcus jannaschii]	54	32	309
385	4	240	479	gi 530878	amino acid feature: N-glycosylation sites, aa 41...43, 46...48, 51...53, 72...74, 107...109, 128...130, 132...134, 158...160, 163...165; amino acid feature: Rod protein domain, aa 169...340; amino acid feature: globular protein domain	54	49	240
7	25	19702	19493	gnl PID e255111	hypothetical protein [Bacillus subtilis]	53	32	210
23	3	2497	2033	gnl PID d102015	[AB001480] SIMILAR TO SALMONELLA TYPHIMURIUM SLVY GENE REQUIRED FOR SURVIVAL IN MACROPHAGE. [Bacillus subtilis]	53	25	465
29	11	9042	10121	gi 143331	alkaline phosphatase regulatory protein [Bacillus subtilis]	53	31	1080
33	3	1479	1009	pir S10655 S106	hypothetical protein X - Pyrococcus woesei (fragment)	53	33	471
36	6	4583	5134	gnl PID e316029	unknown [Mycobacterium tuberculosis]	53	30	552
38	14	8521	8898	gi 580904	homologous to E.coli rnpA [Bacillus subtilis]	53	30	378
52	7	7007	8686	gi 1377831	unknown [Bacillus subtilis]	53	29	1680
54	17	11755	19564	gi 666069	orf2 gene product [Lactobacillus leichmannii]	53	36	2010
56	1	1	681	gi 1592266	restriction modification system S subunit [Methanococcus jannaschii]	53	32	681

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
57	10	9431	8487	gi 1788543	(AE000310) f351: Residues 1-121 are 100 pct identical to YQJL_ECOLI SW: P33944 (122 aa) and aa 152-351 are 100 pct identical to YQJK_ECOLI SW: P33943 (Escherichia coli)	53	31	945
61	1	429	4	gnl pid e236467	B0024.12 [Caenorhabditis elegans]			
71	1	5772	4	gi 393394	TP-291 membrane associated protein [Trypanosoma brucei subgroup]	53	33	426
72	3	894	2840	gi 2293178	(AF008220) YtsD [Bacillus subtilis]	53	33	5769
73	14	9793	9212	gi 1778556	putative cobalamin synthesis protein [Escherichia coli]	53	27	1947
88	7	5217	4342	gi 2098719	putative fimbrial-associated protein (Actinomyces naeslundii)	53	32	582
93	5	2395	1688	gi 563366	glutamate oxidoreductase [Gluconobacter oxydans]	53	38	876
96	9	6632	7762	gi 517204	ORF1, putative 42 kDa protein [Streptococcus pyogenes]	53	33	708
108	8	7629	8600	gi 149581	maturation protein [Lactobacillus casei]	53	42	1131
128	9	6412	6972	gnl pid e317237	unknown [Mycobacterium tuberculosis]	53	32	972
128	12	8429	9253	gi 311070	pentraxin fusion protein [Xenopus laevis]	53	36	561
148	1	3	950	pir a61607 a616	probable hemolysin precursor - Streptococcus agalactiae (strain 74-360)	53	31	825
163	2	2162	3022	gi 1755150	nocturnin [Xenopus laevis]	53	36	948
171	3	2304	2624	gi 1732200	PTS permease for mannose subunit IIPMan [Vibrio furnissii]	53	30	861
182	5	3785	3051	gnl pid d100572	unknown [Bacillus subtilis]	53	32	321
209	3	2948	1935	gi 1778505	ferric enterobactin transport protein [Escherichia coli]	53	35	735
218	5	3884	2406	gi 40162	murE gene product [Bacillus subtilis]	53	28	1014
250	3	473	790	gnl pid e334776	YibH protein [Bacillus subtilis]	53	34	1479
275	1	1	1611	gnl pid d101314	Yqew [Bacillus subtilis]	53	30	318
332	1	544	2	gi 409286	barU [Bacillus subtilis]	53	35	1611
2	2	2543	3445	gnl pid e233879	hypothetical protein [Bacillus subtilis]	53	31	543
3	22	22402	23376	gi 38969	lacF gene product [Agrobacterium radiobacter]	52	39	903
5	3	8094	2356	gnl pid e24915	IqA1 protease [Streptococcus sanguis]	52	36	975
22	26	19961	20212	gi 152901	ORF 3 [Spirochaeta aurantia]	52	32	5739
22	31	23140	24666	gi 289262	comE ORF3 [Bacillus subtilis]	52	35	252
27	6	5397	4801	gi 39573	820 (AA 1-178) [Bacillus licheniformis]	52	32	1527
						52	35	597

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
35	10	8604	7357	gi 508241	putative O-antigen transporter [Escherichia coli]	52	27	1248
45	4	4801	3662	gnl PID d102243	(AB005554) homologs are found in E. coli and H. influenzae; see SWISS_PROT ACC# P42100 (Bacillus subtilis)	52	36	1140
48	18	14385	13726	gnl PID e205174	orf2 (Lactobacillus helveticus)			
49	4	5321	5755	gi 2317740	(AF013987) nitrogen regulatory IIA protein (Vibrio cholerae)	52	25	660
54	4	2773	4668	gi 1500472	M. jannaschii predicted coding region MJ1577 (Methanococcus jannaschii)	52	19	435
54	6	5250	4969	gi 2182453	(AE000079) Y4iO (Rhizobium sp. NCR234)	52	36	1896
66	6	8400	6955	gi 43140	TRK protein [Escherichia coli]	52	40	282
71	26	30659	31312	gnl PID e314993	unknown (Mycobacterium tuberculosis)	52	30	1446
75	2	1673	1035	gnl PID d102271	(AB001683) FarA (Streptomyces sp.)	52	23	654
81	3	1439	2893	gnl PID e311458	rhmannulose kinase (Bacillus subtilis)	52	27	639
81	8	4987	5781	gi 147403	mannose permease subunit II-P-Man [Escherichia coli]	52	32	1455
83	21	20687	21853	gi 143365	phosphoribosyl aminimidazole carboxylase II (PUR-K; ttg start codon) [Bacillus subtilis]	52	37	795
86	6	5785	4592	gi 1276879	Epaf (Streptococcus thermophilus)	52	37	1167
86	20	19390	17861	gi 454844	ORF 3 [Schistosoma mansoni]	52	26	1194
96	13	10540	9659	gi 288299	ORF1 gene product [Bacillus megaterium]	52	26	1530
111	1	2	2026	gi 148309	cytolysin B transport protein [Enterococcus faecalis]	52	33	882
112	2	1457	2167	gi 471234	orf1 [Haemophilus influenzae]	52	27	2025
118	3	2931	2365	bbs 151233	Hip-24 kDa macrophage infectivity potentiator protein [Legionella pneumophila, Philadelphia-1, Peptide, 184 aa] (Legionella pneumophila)	52	33	711
122	9	5646	5951	gi 8214	myosin heavy chain [Drosophila melanogaster]	52	33	567
122	11	6159	6374	gi 434025	dihydroliipoamide acetyltransferase [Pelobacter carbinolicus]	52	36	306
134	6	4880	6313	gi 153733	M protein trans-acting positive regulator [Streptococcus pyogenes]	52	52	216
135	3	1238	2716	gnl PID e245024	unknown (Mycobacterium tuberculosis)	52	43	1434
141	3	1681	2319	gnl PID d100573	unknown [Bacillus subtilis]	52	35	1479
161	4	2562	5024	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)	52	32	639
173	2	968	183	gi 1215693	putative orf; GT9_orf434 [Mycoplasma pneumoniae]	52	36	2463
						52	30	786

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
198	6	4400	3567	gnl pid e313010	hypothetical protein [Bacillus subtilis]	52	26	834
210	12	8844	9107	gi 497647	DNA gyrase subunit B [Mycoplasma genitalium]	52	38	264
214	10	5264	5431	gi 550697	envelope protein [Human immunodeficiency virus type 1]	52	36	168
225	1	15	884	gi 1552773	hypothetical [Escherichia coli]	52	34	870
230	1	39	362	gnl pid d100582	unknown [Bacillus subtilis]	52	28	324
287	1	871	2	gnl pid e335028	protease/peptidase [Mycobacterium leprae]	52	29	870
363	2	1305	4	gi 393394	Tb-291 membrane associated protein [Trypanosoma brucei subgroup]	52	32	1302
23	2	2048	1173	gnl pid e254943	unknown [Mycobacterium tuberculosis]	51	30	876
29	3	742	1521	gi 9229900	5'-methylthioadenosine phosphorylase [Sulfolobus solfataricus]	51	31	780
45	1	410	1597	gi 1877429	integrase [Streptococcus pyogenes phage T12]	51	32	1188
48	26	19227	18946	gi 2314455	[AE000633] transcriptional regulator (tenA) [Helicobacter pylori]	51	33	282
73	5	4276	4016	gi 474177	alpha-D-1,4-glucosidase [Staphylococcus xylosus]	51	31	261
81	11	8935	12057	gi 3111070	pentraxin fusion protein [Xenopus laevis]	51	31	3123
83	5	1195	1986	gnl pid d101316	YqfI [Bacillus subtilis]	51	33	792
98	10	7531	8538	gi 41500	ORF 3 (AA 1-352); 38 kD (put. ftsX) [Escherichia coli]	51	28	1008
113	6	3908	5173	gi 466882	pps1; B1496_C2_189 [Mycobacterium leprae]	51	27	1266
124	1	326	57	gi 2191168	[AF007270] contains similarity to myosin heavy chain [Arabidopsis thaliana]	51	32	270
129	10	7286	6816	gi 1046241	orf14 [Bacteriophage HP1]	51	30	471
143	3	4963	3983	gi 1354935	probable copper-transporting atpase [Escherichia coli]	51	26	981
148	15	11359	10226	gi 2293256	[AF008220] putative hippurate hydrolase [Bacillus subtilis]	51	36	1134
149	8	6003	7313	gi 1633572	Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]	51	21	1311
151	9	12092	11550	gnl pid e281580	hypothetical 40.7 kd protein [Bacillus subtilis]	51	34	543
159	6	2555	3208	gi 146944	CMP-N-acetylneuraminic acid synthetase [Escherichia coli]	51	36	654
174	1	1797	4	gi 1773166	probable copper-transporting atpase [Escherichia coli]	51	28	1794
265	4	2231	1773	gnl pid e256400	anti-P. falciparum antigenic polypeptide [Saimiri sciureus]	51	18	459
277	2	643	1311	pir S32915 S329	plid protein - Neisseria gonorrhoeae	51	33	669

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
350	1	890	3	gi 290509	o307 [Escherichia coli]	51	30	888
363	4	1228	4085	gi 1707247	partial CDS (Caenorhabditis elegans)	51	23	3258
367	1	1701	4	gi 393394	TD-291 membrane associated protein (Trypanosoma brucei subgroup)	51	32	1698
15	5	5174	4497	gnl PID e58151	F3 [Bacillus subtilis]	50	38	678
16	4	2220	2582	gnl PID e325010	hypothetical protein (Bacillus subtilis)	50	29	363
19	5	2591	4159	gi 1552733	similar to voltage-gated chloride channel protein (Escherichia coli)	50	30	1569
25	4	2701	1997	gi 887849	ORF_f219 [Escherichia coli]	50	27	705
35	1	211	417	gnl PID e236697	unknown (Saccharomyces cerevisiae)	50	33	207
39	4	3416	5152	gnl PID d100974	unknown (Bacillus subtilis)	50	27	1737
51	7	4000	5181	gi 1592027	carbamoyl-phosphate synthase, pyrimidine-specific, large subunit [Methanococcus jannaschii]	50	27	1182
51	9	7179	8303	gi 1591847	type I restriction-modification enzyme, S subunit (Methanococcus jannaschii)	50	28	1125
52	8	8740	9534	gi 144297	acetyl esterase (xymC) (Caldococcus saccharolyticum)	50	34	795
52	16	16591	15770	gi 2108229	basic surface protein (Lactobacillus fermentum)	50	34	822
57	7	6031	6336	gi 2275264	60S ribosomal protein L78 (Schizosaccharomyces pombe)	50	40	306
71	23	29348	28383	gnl PID d101328	XqJA [Bacillus subtilis]	50	30	966
86	12	11155	10769	gnl PID e324964	hypothetical protein (Bacillus subtilis)	50	24	387
93	2	1205	330	gi 1066016	similar to Escherichia coli pyruvate, water dikinase, Swiss-Prot Accession Number P23538 (Pyrococcus furiosus)	50	24	876
96	5	1673	2959	gnl PID e322433	gamma-glutamylcysteine synthetase (Brassica juncea)	50	29	1287
98	2	218	1171	gi 151110	leucine-, isoleucine-, and valine-binding protein (Pseudomonas aeruginosa)	50	30	954
103	4	3303	2785	gi 1541330	O-antigen ligase (Salmonella typhimurium)	50	31	519
115	5	6480	5980	gi 895747	putative cel operon regulator (Bacillus subtilis)	50	26	501
129	11	7559	7305	gi 1216475	skeletal muscle ryanodine receptor (Homo sapiens)	50	32	255
129	13	8192	7965	gi 152271	319-kDa protein (Rhizobium meliloti)	50	30	228
151	5	7634	6819	gi 40348	put. resolvase Tnp I (AA 1 - 284) (Bacillus thuringiensis)	50	35	816
153	1	1	597	gnl PID d102015	(AB001488) SIMILAR TO NITROREDUCTASE (Bacillus subtilis)	50	29	597

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
155	5	5986	5432	gi1276880	EpsG (Streptococcus thermophilus)	50	28	555
160	9	7390	6323	gi1786983	[AE000179] o331; 92 pct identical to the 333 aa hypothetical protein YBHE_ECOLI SW: P52697; 26 pct identical (7 gaps) to 167 residues of the 373 aa protein MLE-TRICU SW: P46057; SW: P52697 [Escherichia coli]	50	30	1068
163	6	7396	8091	gnl P1D d101313	Yqen [Bacillus subtilis]			
167	6	5332	3940	gi1413926	ipa-2r gene product [Bacillus subtilis]	50	22	696
169	2	807	130	gnl P1D e304540	endolysin [Bacteriophage Baefille]	50	27	1293
171	5	3168	4025	gi1606080	ORF_o290; Geneplot suggests frameshift linking to o267, not found [Escherichia coli]	50	35	678
210	11	8151	8414	gi1330038	HRV 2 polypeptide [Human rhinovirus]	50	27	858
364	1	1538	135	gi1393396	Tb-292 membrane associated protein [Trypanosoma brucei subgroup]	50	25	264
10	7	5911	5090	gi144859	ORF B [Clostridium parfringens]	50	31	1404
26	5	10754	9768	gi1424440	ATP-dependent nuclease [Bacillus subtilis]	49	24	822
66	7	9777	8398	gi1414170	trkA gene product [Methanococcus marisnigellii]	49	31	987
77	6	5364	4648	gnl P1D e285322	Reck protein [Mycobacterium smegmatis]	49	26	1380
82	13	12689	13249	gnl P1D e255091	hypothetical protein [Bacillus subtilis]	49	28	717
93	9	4866	4531	gi140067	X gene product [Bacillus sphaericus]	49	20	561
112	5	4019	4948	gi1574380	lic-1 operon protein (licB) [Haemophilus influenzae]	49	26	336
129	7	6058	4949	gnl P1D e267587	Unknown [Bacillus subtilis]	49	27	930
135	5	3875	4438	gi139573	P20 (AA 1-178) [Bacillus licheniformis]	49	35	1110
154	2	1423	1953	gnl P1D d101102	regulatory components of sensory transduction system [Synechocystis sp.]	49	25	564
156	5	2878	1637	gnl P1D d101732	hypothetical protein [Synechocystis sp.]	49	29	531
173	5	3500	2940	gi1490324	ORF X gene product (unidentified)	49	25	1242
182	1	1057	2	gi1331002	first methionine codon in the ECLF1 ORF [Salmirine herpesvirus 2]	49	30	561
192	6	5352	3667	gi12394472	[AF024499] contains similarity to homeobox domains [Caenorhabditis elegans]	49	25	1056
253	4	1129	1350	gi1531116	SIR4 protein [Saccharomyces cerevisiae]	49	23	1686
277	1	600	136	gi1396844	ORF (18 kDa) [Vibrio cholerae]	49	23	222
327	3	1435	887	gi1733524	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostellium discoideum]	49	32	965
						49	24	549

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
365	3	1436	132	gi 3933394	TB-291 membrane associated protein (Trypanosoma brucei subgroup)	49	31	1305
33	7	4461	3277	gi 145644	codes for a protein of unknown function (Escherichia coli)	48	26	1185
40	2	652	1776	gnl pid e290649	ornithine decarboxylase (Nicotiana tabacum)	48	29	1125
67	4	1377	2384	gi 1772652	2-keto-3-deoxygluconate kinase (Haloferax alicantei)	48	30	1008
74	2	4269	3871	gi 2182678	(AE000101) Y4WJ (Rhizobium sp. NGR234)	48	27	399
81	2	1326	541	gi 153672	lactose repressor (Streptococcus mutans)	48	33	786
81	4	2981	3646	gi 146042	fucose-1-phosphate aldolase (fucA) (Escherichia coli)	48	30	666
97	1	602	51	gi 153794	rsg (Streptococcus gordonii)	48	29	552
110	1	1	3132	gi 1381114	prtb gene product (Lactobacillus delbrueckii)	48	23	3132
131	5	2914	2147	gnl pid e183811	Acyl-ACP thioesterase (Brassica napus)	48	27	768
133	4	3494	2628	gnl pid e261988	putative ORF (Bacillus subtilis)	48	27	867
139	6	4231	4599	gi 1049388	ZK470.1 gene product (Caenorhabditis elegans)	48	23	369
139	8	5036	5665	gi 1022725	unknown (Staphylococcus haemolyticus)	48	29	630
140	12	11936	11007	gnl pid d102049	H. influenzae, ribosomal protein alanine acetyltransferase; P44305 (189)	48	27	930
146	9	5670	4654	gi 1591731	melvalonate kinase (Methanococcus jannaschii)	48	24	1017
161	3	1280	2374	gnl pid d101578	Collagenase precursor (EC 3.4.-.-). (Escherichia coli)	48	24	1095
172	11	10581	11048	gnl pid d101132	hypothetical protein (Synecocystis sp.)	48	27	468
182	4	2930	2586	gi 40067	X gene product (Bacillus sphaericus)	48	37	345
210	15	10786	11196	sp p13940 LE29_	LATE EMBRYOGENESIS ABUNDANT PROTEIN D-29 (LEA D-29).	48	30	411
214	12	6231	6482	gi 40389	non-toxic components (Clostridium botulinum)	48	26	252
221	1	704	3	gi 1573364	H. influenzae predicted coding region HI0392 (Haemophilus influenzae)	48	27	702
227	2	647	3928	gi 1673693	(AE000005) Mycoplasma pneumoniae, C09_orf718 Protein (Mycoplasma pneumoniae)	48	30	3282
253	2	480	758	gnl pid e236697	unknown (Saccharomyces cerevisiae)	48	31	279
363	3	1874	1122	gi 18137	cgcr-4 product (Chlamydomonas reinhardtii)	48	40	753
389	1	505	2	gi 18137	cgcr-4 product (Chlamydomonas reinhardtii)	48	38	504
3	21	20879	22258	gnl pid e264778	putative maltose-binding protein (Streptomyces coelicolor)	47	33	1380

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins sMilar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
6	4	4089	4658	gi 39573	P20 (AA 1-178) [Bacillus licheniformis]	47	23	570
15	3	3736	1760	gnl PID d100572	unknown [Bacillus subtilis]	47	25	1977
35	15	14516	13263	gi 1773351	Cap5L [Staphylococcus aureus]	47	20	1254
51	6	3547	4002	pir A37024 A370	32K antigen precursor - Mycobacterium tuberculosis	47	38	456
55	8	10154	9273	gi 39848	U3 [Bacillus subtilis]	47	26	882
92	4	1753	3276	gnl PID e280611	PCPC (Streptococcus pneumoniae)	47	35	1524
127	9	5589	5386	gi 1786458	(AE000134) f120; This 120 aa orf is 76 pct identical (0 gaps) to 42 residues of an approx. 48 aa protein Y127_HAEN SW: P43949 [Escherichia coli]	47	32	204
130	2	1232	1759	gnl PID e266555	unknown [Mycobacterium tuberculosis]	47	23	528
140	4	4951	3542	gnl PID d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus [Bacillus subtilis]	47	24	1410
151	4	6814	6200	gi 1522674	M. jannaschii predicted coding region MJEC41 [Methanococcus jannaschii]	47	27	615
157	3	803	1174	gnl PID d101320	YggZ [Bacillus subtilis]	47	25	372
178	5	3267	2155	gi 2367190	(AE000390) o334; sequence change joins ORFs yvjR & yjgS from earlier version (VGJR_ECOLI SW: P42599 and VGJS_ECOLI SW: P42600) [Escherichia coli]	47	30	1113
273	1	2	1549	gnl PID e254973	autolysin sensor kinase [Bacillus subtilis]	47	32	1548
300	2	880	644	gi 1835755	zinc finger protein Png-1 (Mus musculus)	47	22	237
54	14	14182	12638	pir S43609 S436	rofa protein - Streptococcus pyogenes	46	24	1545
88	1	2	1018	gnl PID e223891	xylose repressor (Anaerocellum thermophilum)	46	27	1017
96	7	4553	5860	gnl PID d101652	ORF ID: o34785; similar to [SwissProt Accession Number P45272] [Escherichia coli]	46	23	1308
112	1	1127	3	gi 2209215	(AF004325) putative oligosaccharide repeat unit transporter [Streptococcus pneumoniae]	46	24	1125
122	13	7308	7982	gi 1054776	hr44 gene product (Homo sapiens)	46	34	675
127	14	9198	8125	gi 1469286	afua gene product [Actinobacillus pleuropneumoniae]	46	28	1074
132	4	7093	6197	gi 153794	rpg [Streptococcus gordonii]	46	26	897
140	8	8220	7723	gi 1235795	pullulanase [Thermoanaerobacterium thermophilum]	46	21	498
140	9	9205	8315	gi 407878	leucine rich protein [Streptococcus equisimilis]	46	27	891

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
162	1	1	1125	gi11143209	ORF7: Method: conceptual translation supplied by author (Shigella sonnei)	46	25	1125
199	1	1	585	gi11947171	(AF000299) No definition line found (Caenorhabditis elegans)	46	28	585
223	3	1971	1477	sp102562 MYSS_	MYOSIN HEAVY CHAIN, SKELETAL MUSCLE (FRAGMENTS)	46	27	495
232	2	760	1608	gi11016112	lyc38 gene product (Cyanophora paradoxa)	46	28	849
292	1	687	220	gi11673744	(AE000011) Mycoplasma pneumoniae, cytidine deaminase; similar to GenBank Accession Number C5312, from M. pirum (Mycoplasma pneumoniae)	46	29	468
30	8	5843	6472	gi11788049	(AE000270) o235; This 235 aa orf is 29 pct identical (10 gaps) to 198 residues of an approx. 216 aa protein YTXB_BACSU SW: P06568 (Escherichia coli)	45	24	630
48	6	3461	3868	gi1222339	unknown (Acetobacter xylinum)	45	29	408
60	1	307	2	gi11699079	coded for by C. elegans cDNA yk41h4.3; coded for by C. elegans cDNA yk148g10.5; coded for by C. elegans cDNA yk152g5.5; coded for by C. elegans cDNA yk39a10.5; coded for by C. elegans cDNA cm20g10; coded	45	36	306
72	16	14371	14874	gi11321900	NADH dehydrogenase (ubiquinone) (Artemia franciscana)	45	25	504
99	7	9158	7941	gi1152192	mutation causes a succinylglucan-minus phenotype; ExoQ is a transmembrane protein; third gene of the exoYQ operon; putative (Rhizobium meliloti)	45	28	1218
127	12	7046	6606	bhs1153689	HitB-iron utilization protein (Haemophilus influenzae, type b, DL42, NTHI TN106, Peptide, 506 aa) (Haemophilus influenzae)	45	24	441
137	5	1561	2619	gi1472921	V-type Na-ATPase (Enterococcus hirae)	45	33	1059
209	1	774	364	gi1304141	restriction endonuclease beta subunit (Bacillus coagulans)	45	28	411
314	1	604	2	gi11480457	latex allergen (Hevea brasiliensis)	45	31	603
20	18	19782	20288	gi1433942	ORF (Lactococcus lactis)	44	26	507
87	8	7030	6452	gi1537207	ORF_f277 (Escherichia coli)	44	26	579
166	5	4909	4037	gnl PID a308082	membrane transport protein (Bacillus subtilis)	44	25	873
247	1	818	75	gnl PID d100718	ORF1 (Bacillus sp.)	44	20	744
32	3	1885	3876	gi12351768	Pspa (Streptococcus pneumoniae)	43	24	1992
36	17	15467	18256	gi1045739	M. genitalium predicted coding region NC064 (Mycoplasma genitalium)	43	26	2790
54	15	14656	17343	gi1520541	penicillin-binding proteins 1A and 1B (Bacillus subtilis)	43	27	2688
67	2	696	1352	gi1536934	yjca gene product (Escherichia coli)	43	29	657
139	2	2416	338	gi1396400	similar to eukaryotic Na+/H+ exchangers (Escherichia coli)	43	24	2079

Common proteins

[illegible]

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1	4	3428	3009
1	6	4611	4964
3	2	818	994
3	3	1182	1574
3	7	5382	6497
3	25	25046	25396
3	26	25625	26317
6	2	1519	1689
6	14	12875	12618
6	15	13215	12841
6	18	15977	15390
7	12	9955	9419
7	13	10161	9910
8	6	3915	4280
9	9	6024	5704
10	8	6909	6298
10	9	7136	6888
10	11	7968	7672
12	1	1140	4
12	3	1779	1456
14	2	1913	1434
16	1	1	243
16	5	5675	3087
17	1	324	34
17	3	1451	1050
17	9	4890	4465
20	14	14544	15893

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
21	3	3359	2589
21	5	4802	4482
22	21	17099	17362
22	25	19467	19982
22	33	25540	25764
22	35	26388	26218
22	36	26382	27572
23	7	6655	6032
23	8	7132	6653
24	1	36	518
25	5	3009	2641
27	4	4819	4223
27	5	4789	4956
28	5	3017	1797
28	8	4272	3850
28	10	5028	4597
28	11	5746	5072
29	7	5596	4919
29	8	5039	5518
29	9	5595	8207
30	9	6511	6263
31	6	2664	2344
32	5	5203	5538
33	8	5327	4668
34	10	8024	7740
34	12	9360	8641
34	13	9667	9377

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
34	18	13104	11902
35	11	9688	8588
35	12	11073	9670
36	2	334	1041
36	12	11120	10893
36	13	10993	11388
36	15	12172	14595
38	7	4269	4577
38	8	4480	5001
38	10	5517	5711
38	17	10732	11376
40	3	1728	3143
43	1	172	5
43	7	8884	8732
43	8	9568	9071
44	4	4831	6831
45	3	3204	3665
45	4	3875	3468
46	7	6074	7081
48	5	3196	3582
48	8	4579	4229
48	11	9323	8922
48	16	13042	12494
48	20	16342	15764
48	24	17971	18351
48	30	21979	21776
49	1	209	3

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
50	4	3107	2672
51	5	3239	3598
52	11	12146	12883
54	7	5588	5187
54	8	6013	5459
54	9	6004	6210
54	16	17685	17506
55	9	10515	10123
55	12	11947	12141
56	3	935	1387
56	4	1496	1939
57	3	1624	2130
57	4	2100	2501
58	6	7541	7335
59	1	2	430
59	4	2416	2736
59	5	2734	3063
59	8	4743	5549
59	9	5459	5929
60	6	5741	6451
61	3	2395	1772
61	5	3316	3176
64	1	2722	2
66	2	1180	3147
66	8	9082	9495
67	3	1343	1182
69	2	1165	980

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
70	5	4059	3922
70	6	4215	4057
70	9	5268	5504
71	15	20351	21901
71	16	21859	22338
71	19	26204	27556
72	9	8458	8081
73	4	3815	4216
73	6	4214	4582
73	7	4369	4773
73	10	7183	6428
73	15	9462	9668
76	1	524	195
76	2	867	535
76	11	8602	9210
80	6	7924	8109
81	1	244	2
81	10	6631	8931
83	4	1872	1150
83	17	16810	16460
84	3	4464	2929
86	2	2147	1092
86	4	3606	2875
86	19	16767	17114
87	5	5326	5000
87	7	6459	6001
87	9	7224	7006

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
87	18	17930	17670
87	19	18275	17928
88	2	1619	1840
88	4	2711	2878
88	9	6252	6016
89	3	2634	1621
89	9	7371	6868
90	2	899	2395
90	3	1143	952
91	3	2959	3141
91	4	3170	3691
91	6	4253	4573
93	1	391	2
93	6	2648	2379
93	8	4533	3712
96	1	3	182
96	2	904	632
96	3	1407	1147
96	4	1250	1420
97	9	7043	6753
99	15	18522	18692
99	17	19717	19541
100	2	4094	1980
103	1	48	299
103	6	4924	4373
104	5	6142	6735
105	7	6098	6517

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
106	1	1	363
106	10	9832	10212
108	1	2	268
111	3	3417	3788
111	4	3809	4606
115	10	10854	10438
116	3	2873	2121
118	2	2274	1357
122	4	2698	2333
122	10	5858	6199
122	12	6301	7416
124	2	346	690
128	4	2544	3368
129	1	689	102
129	2	1011	724
129	8	6454	6056
129	9	6540	6277
129	12	7809	7621
131	3	1433	756
131	10	5972	5673
134	11	11838	11209
135	2	625	1140
136	4	2913	3830
137	2	325	134
139	12	14027	14521
139	13	14840	14532
139	14	15363	14875

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
140	20	19822	20838
142	1	1	285
146	3	760	479
146	4	1149	778
146	7	3604	2885
146	13	8223	9401
146	14	9399	10676
146	15	10052	9750
147	7	7488	7276
147	9	8913	8647
148	7	5298	4765
149	1	2	1936
149	3	2557	2880
149	9	6258	6070
150	2	1355	579
150	3	2556	1909
153	3	2061	2642
154	3	1953	1741
155	2	2181	1411
156	8	4550	4311
157	1	37	294
159	2	631	780
159	4	1384	1722
159	7	3271	4017
161	2	1332	1018
165	3	5535	4945
166	6	5406	4972

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
167	9	6075	6395
169	5	2828	3205
170	7	6485	6243
170	8	6964	6362
170	9	7303	6962
170	11	8790	7906
171	9	7150	7476
172	5	2298	1948
173	4	2913	2677
175	2	659	835
175	3	893	1789
176	2	1487	546
176	3	2200	1466
177	9	4686	4925
177	10	4923	5177
177	11	5111	5347
177	13	7396	8703
178	6	3452	3724
181	5	1853	2473
182	2	2112	1102
182	3	2617	2006
183	2	2126	2320
185	5	4683	4219
185	6	4846	4634
187	4	2940	3557
188	4	3686	4363
188	5	4183	4821

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
188	6	5882	6493
189	5	3143	2844
189	9	5956	5564
191	1	618	4
191	11	10357	10001
192	3	2861	2268
192	4	3081	2878
192	7	6800	5331
193	3	997	839
194	4	2315	2127
195	5	6249	4543
195	6	6620	6231
196	2	1553	1849
197	1	1	861
198	9	6844	6644
200	5	5329	5769
200	6	5993	6595
204	5	3914	3276
205	2	447	1709
209	4	2038	2460
209	5	2458	2682
210	10	7370	8230
210	13	9029	10441
210	14	10439	10705
214	5	2581	2330
214	9	5065	5277
214	11	5996	5754

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
217	2	541	194
218	2	914	1432
218	3	1410	1972
218	6	3639	3821
219	1	458	39
220	1	869	600
223	4	2617	1964
227	1	1	510
234	4	1539	1312
234	6	2116	1838
235	1	52	312
235	2	310	687
238	1	660	64
246	1	1	270
248	1	3	362
248	2	443	1222
254	3	2789	792
258	2	1179	1616
260	3	1770	2123
263	1	653	177
263	4	2244	1900
263	5	3569	2973
266	1	1	342
266	2	177	1022
270	2	1124	1681
272	1	857	186
275	2	1684	2295

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
278	1	2	406
282	1	714	391
282	4	1463	1134
287	2	1119	826
288	1	540	4
289	1	684	4
291	5	1589	1858
293	2	2539	2925
294	1	21	608
296	2	494	700
296	3	670	843
302	1	261	530
309	3	559	350
310	2	249	1889
316	2	2087	1818
317	2	1048	584
318	2	313	777
319	3	477	133
327	2	912	607
331	1	1	549
333	1	2	535
333	2	465	82
333	3	127	342
341	1	1	705
345	2	895	701
346	2	750	199
349	1	1	198

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
350	2	81	413
355	1	44	973
358	2	636	448
360	2	948	628
364	2	1639	1265
378	1	345	1004
379	2	683	510
381	1	109	693
385	1	150	4
385	2	269	30

148

(1) GENERAL INFORMATION:

(i) APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
Brian A. Dougherty

(ii) TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

(iii) NUMBER OF SEQUENCES: 391

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.
(B) STREET: 9410 Key West Avenue
(C) CITY: Rockville
(D) STATE: Maryland
(E) COUNTRY: USA
(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
(B) COMPUTER: HP Vectra 486/33
(C) OPERATING SYSTEM: MSDOS version 6.2
(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

149

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Brookes, A. Anders

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PB340P1

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8512

150

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5625 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCAAGCAAAA CCAGCTACAG CTAAAGGAAC TTACGTAACA AACTTGACTA TCACAACCTAC	60
TCAAGGTGTT GGTATCAAAG TTGACGTAAA CTCACCTTAA TCAGTAGTTA AAGTAATGTA	120
AAAAAGTTGA AGACGCTATG TCTCAACTTT TTTTGATGTA CGACGGGCAT GTTGTATAGT	180
AGATGTGTAC TATTCTAGTT TCAATCTACT ATAGTAGCTC AGAAGTCGGT ACTTAAACGT	240
GCTATATCAA AACCAGTCCT TGAAAAACGT GGACTGGTTT CGTGTTTGGA TTATTACCTT	300
GAACGACATG CGTTAAAAGT TAGTTGAACC GCCGTATGCC GAACGGACGT ACGGTGGTGT	360
GAGAGGGGCT AGAGATTATC CCCTACTCGA TTTGCGAAATC TAGTGGAATG AATCTGGAAT	420
AGTCCATCGA GCTTTCTAAT ACTCTTCGAA AATCTCTTCA AACCACGTCA ACGTCGCCCTT	480
GCCGTGCGTA TGGTACTGA CTTGTCAGT TCTATCCACA ACCTCAAAAC AGTGTTTTGA	540
GCTGACTACG TCAGTTCCAT CTACAACCTC AAAACAGTGT TTTGAGCAAC CTGCGGCTAG	600
TTTCCTAGTT TGCTCTTTGG TTTTCATTGA GTATAACACA TTGTTAGAAG TTGTTTAA	660
TTTCCTAATC AGTTTGTTC AATTTACCTT CGATATATTA TATCCCATAG TTAAGGTGG	720
TCATACAGAT GATTATAGTC ATGGAGCCGT AAAACTTAGT GTTCTTTTAG TTGACAAAGA	780
TGCCATGAAA AAAATATTTG TAACTGTAAT AGGATATTTT GAAATAAATA TAGATGAAAA	840
TATCACCGAT ATTCTATACG TAAATGGTAC TGCTATTCTT TATCTTTATT TACGTTCAAT	900
TGTTTCAATA GTTTCGGCAA TTGATAGCAG TGAAGCAATG TTGCTACCTA TCATTAATGT	960
TTTAGAGTTA CTAGATAAAT CTCAACCTTT TGAAGAAGAA TAATTTATTA GCTCACTAAA	1020
TTGAGGGTAA GGAAAAGTAA AAGCAGTAAG AAAAATGTCT TGCATTATAC AGCAACCTTT	1080
TGGGAATGAG TGGATGGATT GAATAAAATT TGATTAAGAG TGGATGATTT ATCTGTAGAT	1140
TATTATTGGA CAGTTAGTCT TGAAGTAGTC TAAGAATTAG GTTATAATCA GTAGAAGCCT	1200
TGCTAATAAT GAGGAGGTTA GTTTATGTAT AGTAGACTGA ATCTAAAATA GTACGAAACA	1260
ATTGCTAAAA CATTATAGA AATTAATTTT ACTTTCCCAA TCGATTTGTT CTCATCTTAT	1320
TTCAATCCGC TATATATTAT GGTATCGAAT CTTATCAGA ATGATAAAAT TAATCAATTG	1380
ATATCTGATT ACAAACAGAA TATGAAAGCT TTTTATATCA CTATTGAAAA ATTTATACGA	1440

151

GATGATGAAA GCCTTAAGTG TTATTTTATA AAGGTTATTT CAAGTCGTC CAAGGTAACA	1500
AGTCTAGATC AGATTGAAGC TGATAAAACG ATACAAAGAA AATATTCAAG TGAGCTAAAA	1560
AAATTTATTG GATTTTATAA TGAGATTATT TGTGAGGAAA ATAGTTTCCT ACATGTACGA	1620
AAGAGGTGGT CGAGTTGCTT TAGGTAGTCG ATGCGTGAGT TGATAATTCT CAGGGTATGG	1680
ACTTCTTTTT CATGAATGAG GTAAAAGAGC AGGTATTGTT TAGAGACAAT CATTCGAGC	1740
ATATTTTCTG GATAGAGGGA GTATCCGATT TTATGATCAA AGTTAATACC GCCCTCTGGT	1800
GAGAAGATGA GTAGGTTGGT AATTTAAACT ATTAACAGA ATTTTGGATT AAAAGTATTA	1860
TTTCATGAGA GAAATCCTAA TTTCACAATC CATAGGCAAA CGCTTGCATT TCGTTTTTTA	1920
TTGGACTATA ATAGGTTGGT ATAAAGCCTT CTGTAGTAAT AAAATGTAGA AGGTGTAGAA	1980
AGTAAGGATT TAGAATATTT GTAGTTAAAA ACACAATGTT GCTATTCCTT ACGATAGGGA	2040
GATAGATATG GCAATGATAG AAGTGAACA TCTTCAGAAA AATTTTGTGA AGACTGTAA	2100
GGAACCGGGC TTGAAGGGG CTTTGCCTC CTTTATTCAT CCTGAAAAGC AGACCTTGA	2160
AGCGGTCAAG GATTTGACCT TTGAGGTCC AAAAGGGCAG ATTTTAGGAT TTATCGGGGC	2220
AAATGGTGCT GGAAGTCGA CAACCATTA AATGCTGACA GGAATTTTGA AACCAACATC	2280
TGGTTTTGT CGGATTAACG GCAAGATTCC CCAGGACAAT CGGCAAGATT ATGTCAAAGA	2340
TATTGGCGTA GTCTTTGGAC AACGCACCCA GCTATGGTGG GATTTGGCTC TGCAAGAGAC	2400
CTACACTGTC TTAAGAGAGA TTTATGATGT GCCAGACTCG CTCTTCATA AGCGTATGGA	2460
CTTTTGAAT GAAGTCTTGG ATTTGAAGGA CTTTATCAAG GATCCCGTGC GGAATCTTTC	2520
ACTGGGACAA CGGATGCGGG CGGATATTGC GGCCTCCTTG CTCCACAATC CCAAGGTTCT	2580
TTTTTTAGAT GAGCCGACCA TTGGTTTGA CGTTTCGGTT AAGGATAATA TTCGTCGGGC	2640
AATTACTCAG ATCAATCAAG AGGAAGAAAC TACCATTCTT TTGACCACTC ACGATTTGAG	2700
TGATATTGAG CAACTTTGTG ATCGGATTTT CATGATTGAC AAGGGGCAAG AGATTTTGA	2760
TGGAACGGTG AGCCAACCTCA AGGAGACCTT TGGTAAGATG AAGACTCTCT CTTTGAAC	2820
GCTACCAGGT CAAAGTCATC TCGTCTCTCA CTATGACGGT CTGTCTGATA TGACCATTGA	2880
TAGACAAGGA AACAGCCTCA ACATTGAATT TGATAGTTCT CGCTACCAGT CAGCTGACAT	2940
TATCAAGCAA ACCCTGTCTG ATTTTGAAAT CCGCGATTG AAGATGGTGG ATACGGATAT	3000
TGAGGATATT ATCCGTCGCT TCTACCGAAA GGAGCTCTAG GATGATCAAA TTGTGGAGAC	3060
GTTATAAACC CTTTATCAAT GCAGGGGTT AGGAGTTGAT TACTTACCGA GTCAACTTTA	3120
TTCTCTATCG GATTGGCGAT GTCATGGGG CTTTGTGGC CTTTATCTC TGGAAGGCTG	3180

152

TCTTTGATTC TTGCAAGAG TCTTTGATTC AGGGCTTCAG TATGGCGGAT ATCACCCTCT	3240
ACATCATCAT GAGTTTGTG ACCAATCTTC TGA TAGATC CGATTTCGTCC TTTATGATTG	3300
GGGAGGAGGT CAAGGATGGC TCCATTATCA TGC GTTTGTT GCGACCAAGTG CATTTTGCGG	3360
CCTCCTATCT TTTCACCGAG CTTGGTTCCA AGTGGTTGAT TTTTATCAGC GTTGGCCTTC	3420
CATTTTAAAG TGTCATTGTC TTGATGAAAA TCATATCGGG TCAAGGTATT GTAGAGGTGC	3480
TAGGATTAAC TGTCATTTAT CTTTTAGCT TAACGCTCGC CTATCTGATT AACTTTTTCT	3540
TTAATATTG CTTTGGATTT TCAGCCTTG TGT TAAAAA TCTTTGGGT TCCAACCTAC	3600
TTAAGACTTC CATAGTGGCT TTTATGTCGG GGAGTTTGAT TCCCTTGGCA TTTTTCCAA	3660
AGGTGTTTC AGATATTCTC TCCTTTTGC CTTTTTCATC CTTGATTAT ACTCCAGTTA	3720
TGATCATGT TGGAAAATAC GATGCCAGTC AGATTCTCA GGCACCTCCT TTGCAGTTCT	3780
TCTGGCTCTT AGTGATGGTG GGATTGTCTC AGTTAATTG GAAACGGGTC CAGTCCTTTA	3840
TCACCATTCA AGGAGGTTAG TATGAAAAA TATCAACGAA TGCATCTGAT TTTATCAGA	3900
CAATACATCA AACAAATCAT GGAATATAAG GTAGATTTG TGGTTGGTGT CTTGGGAGTC	3960
TTTCTGACTC AAGGCTTGAA TCTCTTGTTC CTCAATGTCA TCTTTCAACA TATTCATTTC	4020
CTAGAAGGCT GGACCTTTCA AGAGATAGCT TTCATTTATG GATTTTCCTT GATTCCTAAG	4080
GGAATGGACC ATCTCTTTT TGACAATCTC TGGGCACTAG GGCAACGCCT AGTCCGAAAA	4140
GGGGAGTTG ACAAGTATCT GACTCGTCCC ATCAATCCTC TCTTTCACAT CCTAGTTGAA	4200
ACCTTTCAGA TTGATGCCTT GGGTGAATC TTAGTCGGTG GTATTTTATT GGAACAACA	4260
GTGACCAGCA TTGTTTGGAC TCTTCCAAAA TTCCTGCTTT TCCTAGTTG TATTCCTTTT	4320
GCGACCTGA TTTATACTTC TCTTAAAAATC GCAACAGCCA GTATCGCCTT TTGGACTAAG	4380
CAGTCAGGCG CCATGATTTA CATCTCTAT ATGTTCAATG ACTTTGCTAA GTATCCGATT	4440
TCTATTTACA ATTCTCTTCT TCGTTGGTTG ATTAGCTTTA TCGTGCCTTT CGCCTTTACA	4500
GCCTACTATC CAGCTAGCTA TTTCTTACAG GAAAAGGATG TGTCTTTAA CGTAGGAGGT	4560
TTGATGTTGA TTTCTCTGGT TTTCTTTGTT ATTCCCTTA AACTTTGGGA TAAGGGCTTA	4620
GATTCCTACG AAAGTCCGGG TTCGTAAAAG CTAAAGTAAG ACTAAAATCA AGAAAGAAC	4680
TTATGATGTT TGTAAATGAA GAAGTCAAGG ATGAAAATCA AAAAAAGGCA GTTGTCTGCTG	4740
AGGTTTTGAA GGATTTGCCA GAATGGTTG GAATCCCAGA AAGCACACAA GCCTATATAG	4800
AAGGAACCAC GACACTGCAA GTTTGGACCG CCTATCAGGA GAGTGATTG ACTAGATTG	4860
TAAGCTTATC CTATTCGAGT GAAGATTGTG CAGAGATTGA TTGTCTCGGC GTAAAAAGC	4920
TTATCAAGGT AGAAAAATTG GGAGCCAATT GCTTGCTACT TTAGAGAGTG AAGCTCGTAA	4980

153

AAAAGTTGGT TATCTGCAGG TCAAAACAGT GGCAGAAGGT TCTAATAAAG ATTATGATCG	5040
AACAAATGAC TTTTATCGAG GTCTTGGCTT TAAAAAGTTA GAGATTTTTC CTCAACTATG	5100
GAATCCGCAA AATCCTTGTC AGATTTTGAT TAAAAAGCTT GAATAATATT ACTTGACATC	5160
TATTCTCAGA GTGCTATACT GTAAGTGTA TCGCCGATTT AGCTTAGTTG GTAGAGCAAG	5220
GCACTCGTAA AGCCTAGGTT ATAGGTAGAT AAACGACTGA GGATTGAAA AAATAGATAG	5280
GTAGAAGATA ACCGTTAAGC CTTACTCTTA GCGGTTATTT ATATTGTTTA ATAGCGCTAA	5340
TATTTTATCA ATTATGCCTG TTTTCGTGTT TCTGGTAGTT GTTCAAGTTT ATTGCTACTA	5400
TTTTTGATGG TATGAATGTG CTTATAATGT ATCCCGGTTA ACGAAAGTTT TGGACTTATA	5460
CTCTTCGAAA ATCTCTTCAA ACCACGTCAA CGTCGCCTTG CCGTGCCTAT GGTATGACT	5520
TCGTCACTTC TATCCACAAC CTCAAAACAG TGTTTGTAGT GACTACGTCA GTTCCATCTA	5580
CAACCTCAAA ACACTGTTTT GCCCAATCTG CGGCTAGTTT CCTAG	5625

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7571 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CTCTCCAGCT TTCCTTGCGA GTTGGCCATG TTGTGTCTTT AAGAAGTCTA AAAATATCTC	60
CAATAAAACG CATCGCTCTC TCCTATCTCG TTTCTCTGTG TGTAAGTAC TTGCCACAAT	120
GCTTACAAA TTTATTTACT TCTAGTCGTG TAGGCTTGAG GTTCCGCTG ATCTTGATTG	180
AATAGTTTCT CGAACCACAA ACCGCACAAG CTAGGCTTGC TTTTTTTAGT GCCATAACGC	240
CTCCATCTTA TCCATTATAA CAAGAAAGCT AGGCTTTGAC AAGCATCTTA GCGAAATAGA	300
TTGACTATCG AATCCCATAT TGTTTGAGCC TTTTCCTTAA TCTTCGCATC TGAGATAGCC	360
CGGCTAGCCT CATCTACTAG ACTTTGCGCA CGCCCTCGAA TATCAGACAA ATTATCATCT	420
GTCTGGCTAT TATCATTTGGT TTGTACTTGT CTTTTTGTAT TGGCTGGTGC AATTCCATTT	480
TGCTTATAAG CATTTTCAAC CGTAAAGGTA CTTCTGGCG TATAAGGTAA AATGGTATTG	540
GCAATGTTTC TAAAGACATG AGCTGCACCG TTTGAAGTAG AGCCAGCTAG ATAGTGGTTT	600
TCATCAGTGG TCGGAAAGCC AAGCCAGTGG CTAATCACTA CATCCGGAGT ATAACCAATT	660
ACCCACTGGT CACTTGTGTA CTCCGGATTG AAAACTGCTT CAGTTGTTCC AGTTTTCCTT	720

154

GCCATGACAT AGTCTGCAGG CGATGAACTA ATACCGGTAC CGTTGGTGAA AGTCCCCAAC	780
ATCATACTGG TCATCTTGTC AGCTACAGAC TTATCAATCA CCCGTTTTTG TGAATTTTTA	840
TGACTCGCAA TAACTTGTC ACTAGCATTT TCAATTCTAC TAATAAAATG AGCTTCAGGC	900
ATTAAACCTT CATTTGCAA GCGGCGTAT GCTTGAGCCA TTTGAAGAGG GTTGGTTTCA	960
ACACCGCTTC CCAAGGCGAC ACCAAGAACA CGGTCGACCT TTTCCATGTT GAGTCCGAAT	1020
TTTTCGCCTG CCTCAAAAGC CTTGTCGACA CCCAAATCAT TAACAGTGGC AACAGCAGGT	1080
AGATTAAGCG ATTCTGCCAA GGCTTGATAC ATAGGAACCT CTCGACTCGT TTTGATCCCT	1140
GCATAGTTAT CAACCTTATA GCTGTCATAC TGCATGGTAT GGTATCCAA CTGCTTATTC	1200
AAAGCCCAGC TTGCTTCAAC TGCTGGCGTA TAAACAACCTA AAGGCTTAAT TGTAAGACCA	1260
GGACTACGCT TTGATTGGGT TGCATAGTTG AAATTCGGGA ATCCAGTTTT ATCATTGTCA	1320
GCAACTTGAC CGACAACCTC ACGAACTCCC CCTGTTTTCG GTTCGAGGGC TACACTTCCT	1380
GATTGAGCAA ACGTTCCATC CTCTGCCCTC GGAAATAGCG ATGTGTTTTT ATAAACAATC	1440
TGCATATTTG CTTGGTAGTT TTGGTCCAGC TCTGTGTAAA TGCGGTAGCC ATTATTGACA	1500
ATCTCTTCCT CTGTTAGATT ATACTTGGAA ACAGCTTCAT TAACCACCGC ATCAAAATAA	1560
GAGGGGTAAC GGTAATCTGA GATTTTTCTT TCATACTTAT CGTGCAATTG CGAAGTCATA	1620
TCAACTTCAG CAGCTTTGGT TTCTTGTTTT TTATCAATAT ATCCTGCTGC AACCATATTC	1680
TGCAAGACAG TATCGCGCCG ATTAGTAGAA TCTTCTACGG AATTCAAGGG ATTATACAGT	1740
TCCGGCCCCCT TGAGCATCCC TGCCAGAGTC GCAGCTTGAT CCAGACTCAC TTCTGATGCA	1800
GAAACTCCAA AGTATTTCTT ACTCGCATCT TCTACACCCC ACACACCATT TCCAAAATAA	1860
GCGTTGTAA GGTACATGGT TAGAATTTGC TCCTTACTAT ATTTTTTGCT TAATTCTAAG	1920
GCAAGGAAAA ATTCTTTTCG TTTTCTCTCA ACAGTTTGAT CCTGCGATAA ATAGGCGTTT	1980
TTAGCCAGCT GTTGGGTAAT GGTAGAGCCA CCACCTGAAC GTCCAGCAGT GACAATAGCC	2040
AAGAAAAAAC GGCCATAGTT AATCCCGTCA TTTTATAGA AAGAACGGTC TTCTGTCGCA	2100
ATAACAGCAT TCTGCAAGTT TTTACTGATG TCAGTCAGCT CAACATAGGT TCCCTTTTGA	2160
CCAGACAAGG CACCAGCCTC TTTTCTTCA CGGTCAAAAA TAAGAGTCCG AGTTTTCAAG	2220
GCATTTTGCA AATCATTGAC ATTGGTCGAC TTGGCTACAG CAAACAAATA GATTCCAAC	2280
AGCAAGCCTG CACTCAAACC TAGTATAAGG ATAATCTTG TTAGATGATA ACGACGCCAG	2340
AATTTTCGAA TCGGACCTAC TTGGGCTAAT TTTTTCGAT CACTACGAGA GCGACGTAAG	2400
ATAGTAGAAT CAGAGTCCTC TAGTTCACCT GTTTCTTTT TAAAAAGAGA AAGAAATTC	2460
TCAAATAATT TATCTAATTT CATGCGTTTA TTTTATCATC TTCATCATAG GAAGACAAGA	2520

ATTTAGCTAT TTCCTATCCA AATAGGGCTT TTTTGTGTTAC AATATCTGTA TGCAATTCAC	2580
ATTTACATTA CCCGCCTCTC TACCTCAAAT GACAGTAAAG CAATTACTTG AGGAACAAC	2640
CCTCATCCCT AGAAAAATCC GTCATTTTGT GAGAATCAAG AAACATATTT TGATAAATCA	2700
AGAAGAAGTC CACTGGAAGG AAATCGTAAA TCCTGGAGAT GTTTGCCAGT TGAATTTTGA	2760
CGAGGAAGAT TATTCCTAAA AGACGATCCC TTGGGGCAAC CCAGACTTAG TGCAGGAAGT	2820
TTATCAAGAT CAACACTTGA TTATTGTAAA CAAACCAGAG GGGATGAAAA CGCATGGTAA	2880
TCAACCAAAC GAAATGCCCC TTCTTAACCA TGTCAGTACC TATGTTGGCC AAACCTGCTA	2940
TGTCGTTTAT CGTCTGACA TGGAAACCAG TGGCTTAGTT CTCTTTGCCA AAAATCCTTT	3000
TATCCTGCCC ATTCTCAATC GCTTATTGGA GAAAAAGAG ATTTCTAGAG AATATTGGGC	3060
TCTAGTTGAT GGAAATATCA ACAGAAAAGA ACTTGTTTTC AGAGACAAAA TTGGACGTGA	3120
TCGCCATGAT CGTAGAAAA GAATAGTTGA TGCAAAAAAT GGGCAATATG CTGAAACGCA	3180
TGTAAGCAGA TTAAAGCAAT TCTCAAACAA GACTTCCTTG GCTCATTGCA AGCTAAAGAC	3240
AGGGCGAACC CATCAGATTC GTGTGCACCT TTCGCATCAT AATCTTCCTA TCCTGGGAGA	3300
CCCTCTCTAT AATAGTAAAT CAAAGACAAG CCGGCTTATG CTTCATGCCT TCCGACTTTC	3360
CTTTACCCAC CCACTTACTT TAGAGAAGCT AACTTTCACT ACCCTTTCAA ATACATTTGA	3420
AAAAGAATTA AAAAGAATG GATGATCGTG TCATCCATT TCCCATATAA AAAAGCAAGA	3480
CCACAAAGCC TTGCTTTCTA TCAACTCAAG AATTATTTAG CAATTTTTCG GAAGTATTCA	3540
AGAGTACGAA CAAGTTGTGC AGTGTATGAC ATTCGTTGT CGTACCATGA TACAACTTTA	3600
ACCAATTGTT TACCGTCAAC GTCAGAAGT TTAGTTTGAG TTGCGTCAAA CAATGAACCG	3660
TAAGACATAC CTACGATATC TGAAGATACG ATTGGATCTT CTGTGTAACC GTATGATTCTG	3720
TTTGAAGCTG CTTTCATAGC TGCGTTCCT TCATCAACAG TAACGTTCTT TTCAAGAAGT	3780
GCTACCAATT CAGTAACTGA TCCAGTTGGA GTTGAACGC GTTGTGCAGA TCCGTCAAGT	3840
TTACCATTCA ATTCTGGGAT TACAAGACCG ATAGCTTTTG CAGCACCAGT TGAGTTAGGA	3900
ACGATGTTTG CAGCACCAGC GCGAGCACGG CGAAGGTCAC CACCACGGTG TGGTCCGTCA	3960
AGGATCATTT GGTCAACAGT GTAAGCGTGG ATAGTAGTCA TCAATCCCTC AACAAACCCA	4020
AAGTTGTCTT GAAGAGCTTT AGCCATTGGA GCCAAGCAGT TTGTAGTACA TGAAGCACCT	4080
GAGATAACTG TTTCAGTACC GTCAAGAACG TCGTGGTTAG TGTGAATAC AACTGTTTTA	4140
ACGTCGTTTC CACCAGGAGC AGTGATAACA ACTTTTTTAG CTCCACCTTT AAGGTGTTTT	4200
TCAGCTGCTT CTTTCTTAGC AAAGAAACCA GTAGCTTCAA GAACGATTTC TACACCGTCA	4260

156

GTAGCCCACT CGATTTGTTC TGGATCACGT TCAGCAGAAA CTTTGATGAA TTTACCGTTA	4320
ACTTCAAATC CACCTTCTTT AACTTCAACA GTACCGTCGA AACGACCTTG AGTTGTGTCG	4380
TATTTCAACA AGTGTGCAAG CATAACTGGA TCTGTAAGGT CGTTGATGCG TGTAACCTCA	4440
ACACCTTCTA CGTTTGGAT ACGACGAAA GCAAGACGAC CGATACGTCC GAAACCGTTA	4500
ATACCAACTT TAACTACCAT TAGTGATTTC CTCCTTATGA AAATCATGAA ATTTTATTG	4560
TGAAAAGAGT AACTTGAATC ACTACAAATC ACCTTTCAAC AAACCTATTA TACAACATTT	4620
TGAGTTGAAT TGCAAGTATG GCCATTGTTT TTCTATGTTA GTTCTCTTTT AAGACTGTAA	4680
ACCAAGGAAT CCCTTACTAT TCATAGCATA ACGATTCTAT AGGATCCATT TTAATAATCT	4740
TACGCGCCGG GAAGTAGGCT GAGACATAAC CAAGTAATAG AGCGAAAAC AGAGTTCCTA	4800
AAACAGATAA AAGATTAAAT TTAATAAACCT TAGTGATGGA TGGGTAAAAG TGACTTACAA	4860
TCGCATTGCG CAAACTTCCC ACCCCTTGTG CAACCAAAA TGCCAGCAGC AAGGCGATGC	4920
CTACAATCCA GATAGCCTCG TAAATAAAAA TTCCTTTGAC ATCACGATTC TGATAACCAA	4980
CTGCTTTCAT GACACCTATT TCCTTGGAAC GTTGATGAT ATTGATGTAA ATAATGATAC	5040
CAATCATAAC CGCTGCTACC ACAATAGCTT GTGATGAAAG CACAATCAAT AATCCCTGAA	5100
TAACACGAAT AAAGGTAATC ACAATATCAA GAACTCTCTG TTGAGAAAGC ACAGTATACT	5160
TCTTATTTT CTGTAATTCT TCTGTTACTA CTTTGTCTG TGATGGATCT TTGAGTTCCA	5220
AGATAAAATA AGATACAGCT TTCGTAAATC CAGCCTCTTT CAAAATCGTT TCCATTGTAT	5280
GAGACAGCAT GAAACTGTTG CTGTCCTCCA TGTCATCTTC ATCATGTGAT ACACGTACAA	5340
TCTTCGTTTG AAATGAGCA ATCTTACTAG TTTCCGCAGC ACTTCTTACA ATGCTGGCTG	5400
AGACTGATTT GCCAATAAGA TCATTAGCTG TCAAATTTT TCCTGTCTGT TCATTCCAAT	5460
TTTTTAGTAA ACTGCTTGA ATCGTTAATC CCTGTTTATT TGTATCAGTA TAGAGGGATC	5520
CAGCCAACAC TTTGTCCGTC TCATTATTAC TAACAGAGAT ACTTGTATCA TCATAAAGAC	5580
TCACTACTTG AGCATAAGAA GGCATCGTTT GACTCAGATC CATTCTTGC CCATCTATAG	5640
TAATATTTGA CATGTTTATC CCAAAGGAC TCTCCAAATA TTTAATAGCT TCTTTCCCAA	5700
CTGTATCCGT GATATATAGT CAATTGAAAC AAGAGCAGGA TAAAAAGCC TCGTAAAGG	5760
TATTGCAACT TGGTAATACC TTTTGAGGT GCTTTTGAT ATGAGCCCAT GTTTCTCAA	5820
TAGGATTGTA CTCAGGCGAG TAGGGAGGAA GAGGTAAAAG TTTATGCCCA AACTCTTCGC	5880
ATAAAAGTTC TAGCTTCCCC ATTCTATGGA ATCTTACATT ATCCATAATA ATAACCGATG	5940
GTGTGTTTAA TGTGGTAAG AGAAAATCT GAAACCAAGC TTCAAAAAAG TCGCTCGTCA	6000
TCGTCTCTTC GTAAGTCATT GGAGCGATTA ATTCACCATT TGTTAGACCT GCAACCAAAG	6060

157

AAATCCTCTG ATATCTTCTT CCAGATACTT TGCCTCTTAT TAATTGACCT TTTAATGAGC	6120
GACCATATTC TCGATAAAAA TAAGTATCGA ATCCTGTTTC GTCAATCTAA ACAGGTGCTA	6180
GGTGCTTTAA ACTATTAAAA TTCTTAAGAA ATAAGGCTAC TTTTCTGGG TCTTGTTTCAT	6240
AGTAGGTGTG GTTCTTTTTT CGAGTGTAGC CCATAGCTTT GAGCGTATAG TGGATGGTAG	6300
TTGGATGACA GCCAAATTCA GAAGCTATTT CACTCAAATA AGCGTCTGGA TTGTCAGTAA	6360
GATAGTTTTT AAGTCTATCT CTATCAACCT TTCTTGGTTT TATTCCTTTT ACTTGGTGGT	6420
TTAGCTCTCC TGTTTTCTCT TTTAGCTTTA ACCAGCCATA AATGGTATTA CGTGAGATTT	6480
GGAAAACTG TGATGCTTCT GTTATACTAC CTGTTGCTC ACAATAAGAG AGAACTTTTT	6540
TACGAAAAATC TATTGAATAT GCCATAAAAA GATTATACCA CATTGTGTAC TATTTTGGT	6600
TCATTTTACT ATATTGAAG AGGCGTTTAA ACTATCTGAC ATAAAACTCG TTCTAGAGGA	6660
AAGACATCCT TAAAAAGTT AGTTTATTTT ACAACTTAGA CATCAAGGTA GGTAAACCCC	6720
TTTCATGAAA AATCAAGACT CTTAGCACTA TGGGTTAAAC TACCACTGGA GACGTAATCA	6780
ATCGCTAAAC CACGAAAACG GCTAATAGTG GTCATATCAA TATTTCCAGA ACATTCAATC	6840
CGAGAAGCTC CTGCAATTAG GGTAATGGCC TGTCAATCT GTTCCAATGA CATATTATCC	6900
AACATGATAA TATCAGCACC CGCCGCCGCA GCTTCTTCGG CAGCAGCAAG GCTTCCACT	6960
TCCACCTCGA CCATTTTCAC AAAAGGGGCA TAGGCACGCG CTTGAGCAAT TGCCTTTTGA	7020
ACACTACCTA CTGCCGCAAT GTGATTGTCT TTTAGCAGGA TAGCATCTGA TAAATTAAAG	7080
CGATGATTAT AGCCACCGCC AACTCTCAGC GCATATTTCT CAAAAGACG TAAATTAGGA	7140
GTAGTTTTTC GAGTATCAAA TACCTTAATG CAATCATCGC CTAAGGCTTC TACATAAGCA	7200
GCTGTCATCG AAGCAATCCC TGATAAATGT TGTA AAAAAT TCAAGGCAAC GCGTTCACAT	7260
GTAAAGAGAC TTCTACCGA GCCTATGATT TCTAAAACCA AATCGCCACT AGTCAAACGA	7320
TCCCCATCCT TAAATTGATG AGGATTCTGG AAGGTCACCT CGGCATCAAA TAGGGTAAAA	7380
ACCTTTTGAA AAACGGTTAG CCCCCTAAA ACACCAGCTT CCTTGGCAAA AAGCGACACC	7440
TTGGCTTGGC CATGATGATC AAAAATGGCA TTGGTACTGT AATCTTCGGA ATGAACATCT	7500
TCTCGCAAGG CTGCTTTCAA TGTATCATCT ATTTGAAAAG GGGTTAAATC AGTTGAAATG	7560
ATTGACATCA C	7571

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26385 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTTGCTAGTG GCTTAAATTC TTCAGGAAAA TCAGGCGTAT CTAAAAGTCG TGTCGTTTTT	60
GTTTCATCTA TATAAAGACT TCCTGCTCCC CCTACAATA GAAACGTGT CTGTGTTCCA	120
GCAAGAAGCT GATTAAATAG TTCGATTGAT TTGCTGTGGA GCGGTAGCGT ATCTGGTGTA	180
TAAGCACCAA ACGCTGAAAT AACAGCATCA AATCCAGTAA GATCATCTTT TGTCAACTCA	240
AATAAATCTT TTTAATAAT AGACTCAGCT TGACTTTTGT TTTCAGAACG AACAAATAGCC	300
GTTACTTCAT GTCCTCGTTT GACTGCTTCT TCAACAATG CTTTCCCCGC TTGTCCATTT	360
GCTGCAATAA CTGCTAGTTT CATTTTTTAT ACCTCTCTTG TTGTAATTAT TTTAGTTACA	420
GAAATTGTGA CACTCTTAAT AATCAATGTC AATAGTCTTG CTTAATTATT ATCAAAATAT	480
TTCTACCAAG AAAACTAACC ATGATTCTAG TGAAAAAAA TCTTCTTTGT CAACAAATTT	540
ACTTCTTGT TTTAAACATG CTATAATAAT CATAGCAAGA GATCTAAGTT GTCTGTTTTT	600
TTAAACGAG GTGATTATCA TGCCTAGATT CTATTCCCAT CTCCCCTACT ATCTGGTCAT	660
ATTATTCTTT TATTGGCCAC TTTATGAGTT GTTCTTACTA GTTGTPTCTG ACCCCCTTAC	720
ACTCAAGGGA CTCTATATAA ACAATCTTCT CTTCTTTACA CCTCTGGTAA TCTTGATTGT	780
ATCGTTACTC TATAGCTACC GTTCCGTTT CTCACCTTGA TGGTTAGTTG GTAACGGACT	840
GCTCTTTTAC TTTACTATCA TAACCTTTGG TGAGTTTATA CTAATTACT TGCTAATCTA	900
TGAAACAGTT GCTCTGGTCG GCATGGATTC TGGTATTAGC ATCAAGCATA TTCTACAAAA	960
AATGAAAAAC AAAAACTTT CACAAAATCC TTGAAAAATC TCACAATCAT GCTATAATAA	1020
TCCATAGAGA CAAGTCACCT AGTCCCTTTC TACTAGAGAG TGCCTGGTTG CTGGAAACGC	1080
ATAGGAACTC TAACTGATA CTAATCTTGA GTTTTTATG AAAACATAAA ACGGTGGCCA	1140
CGTTAGAGCC GATCAGAGGT GTCCCTCTCT TTTGAGGTAC ATAAATGAAG GTGGAACCA	1200
GTTGCGACGT CCTTTCGAGG ATGTCGCATT TTTTATTAG GATACTAATT ATGGAGTTGC	1260
AAGAAATTAGT GGAGCGCAGT TGGGCAATCC GACAAGCTTA TCACGAACGT GAAGTTAAGC	1320
ATCATGATTC CAAGTGGACG GTAGAAGAAG ACCTCTTGGC TTTATCTAAT GATATTGGAA	1380
ATTTCCAACG ACTGGTGATG ACAAAGCAAG GACGCTACTA TGATGAAACA CCCTACACAC	1440
TGGAACAAAA ACTTTCAGAA AATATCTGGT GGCTATTAGA ACTTCTCAA CGTTTGGATA	1500
TAGACATTCT GACGGAAATG GAAAACTTC TCTCTGATAA AGAAAAGCAA TTGAACGTTA	1560
GGACTTGGA GTAGTCTGCT GATAAAAAAT CAATGCTTAG AAATATGAA ATAATAAAAA	1620

159

AGGAGAACAT CATGATTAAC ATTACTTTCC CAGATGGCGC TGTTCGTGAA TTCGAATCTC	1680
GCGTAACAAC TTTTGAAATT GCCCAATCTA TCAGCAATTC CCTAGCTAAA AAAGCCTTGG	1740
CTGGTAAATT CAACGGCAAA CTCATCGACA CTACTCGCGC TATCACTGAA GATGGAAGCA	1800
TCGAAATGT GACACCTGAT CACGAAGATG CCCTTCCAAT CTTGCGTCAC TCAGCAGCTC	1860
ACTTGTTTCG CCAAGCAGCT CGTCGTCTTT TCCCAGACAT TCACTTGGGA GTTGGTCCAG	1920
CCATCGAAGA TGGTTTCTAC TACGATACTG ACAACACAGC TGGTCAAATC TCTAACGAAG	1980
ACCTTCCTCG TATCGAAGAA GAAATGCAAA AAATCGTCAA AGAAAACCTC CCATCTATTC	2040
GTGAAGAAGT GACTAAAGAC GAGGCACGTG AAATCTTCAA AAATGACCCT TACAAGTTGG	2100
AATTGATGA AGAACTCA GAAGACGAAG GCGGTTTGAC TATCTATCGT CAGGGTGAAT	2160
ATGTAGACCT CTGCCGTGGA CCTCACGTTT CATCAACAGG TCGTATCCAA ATCTTCCACC	2220
TTCTCCATGT AGCTGGTGG TACTGGCGTG GAAACAGCGA CAACGCTATG ATGCAACGTA	2280
TCTACGGTAC AGCTTGGTTT GACAAGAAAG ACTTGAAAAA CTACCTCAA ATGCGTGAAG	2340
AAGCTAAGGA ACGTGACCAC CGTAACTTG GTAAAGAGCT TGACCTCTTT ATGATTTCAC	2400
AAGAAGTGGG ACAAGGTTTG CCATTCTGGT TGCCAAATGG TCGGACTATC CGTCGTGAAT	2460
TGGAACGCTA CATCGTAAAC AAAGAGTTGG TTTCTGGCTA CCAACACGTC TACACTCCAC	2520
CACTTGCTTC TGTGAGCTT TACAAGACTT CTGGTCACTG GGATCATTAC CAAGAAGACA	2580
TGTTCCCAAC CATGGACATG GGTGACGGG AAGAATTTGT CCTTCGTCCA ATGAATGTC	2640
CGCACCACAT CCAAGTTTC AACACCATG TTCACTCTTA CCGTGAATG CCAATCCGTA	2700
TCGCTGAAAT CGGTATGATG CACCGTTACG AAAAATCTGG TGCCCTCACT GGCCTTCAAC	2760
GTGTACGTGA AATGTCACTC AACGACGGTC ACCTATTCGT TACTCCAGAA CAAATCCAAG	2820
AAGAATTTCA ACGTGCCCTT CAGTTGATTA TCGATGTTTA TGAAGACTTC AACTTGACTG	2880
ACTACCGCTT CCGCTCTCT CTTCGTGACC CTCAAGATAC TCATAAGTAC TTTGATAACG	2940
ATGAGATGTG GGAAAATGCC CAAACCATG TTCGTGCAGC TCTTGATGAA ATGGGCGTGG	3000
ACTACTTTGA AGCCGAAGGT GAAGCAGCCT TCTACGGACC AAAATTGGAT ATCCAGATTA	3060
AAACTGCCCT TGGAAAAGAA GAAACCTTT CTACTATCCA ACTTGATTTC TTGTTGCCAG	3120
AACGCTTCGA CCTCAAATAC ATCGGAGCTG ATGGCGAAGA TCACCGTCCA GTCATGATCC	3180
ACCGTGGGGT TATCTCAACT ATGGAACGCT TCACAGCTAT CTTGATTGAG AACTACAAGG	3240
GGGCCTTCCC AACATGGCTG GCACCACACC AAGTAACCCT CATCCCAGTA TCTAACGAAA	3300
AACACGTGGA CTACGCTTGG GAAGTGGCCA AGAACTCCG TGACCGCGGT GTCCGTGCAG	3360

160

ACGTAGATGA GCGCAATGAA AAAATGCAGT TCAAGATCCG TGCTTCACAA ACCAGCAAGA	3420
TTCCCTTACCA ATTAATTGTT GGAGACAAAG AAATGGAAGA CGAAACAGTC AACGTTTCGTC	3480
GCTACGGCCA AAAAGAAACA CAAACTGTCT CAGTTGATAA TTTTGTTCAA GCTATCCTAG	3540
CTGATATCGC CAACAAATCA CGCGTTGAGA AATAAGAGTC TAGCATAAAA GCCTCCAATC	3600
TGGAGGCTTT TTCTCATCTA TTTTACTCA AGGACTAAGT TCACTTGAGC AAAGTGAATC	3660
CGCACTGTCTG TTCCTTTTCC GACCTCAGAC TCGATACGAA TCTGGTGCCC CAGTTCTTCA	3720
GAAATTTTCT TAGATAGATA AAGGCCAAGT CCAGAGGACT GCTGGGTCAA ACGGCCATTG	3780
TATCCTGAAA AGCCACGTC AAATACTCGG AGGACATCAC TGTTTTTTAT CCGATTCCC	3840
GTATCTTTGA TACAAAGCTC TTGGTCATCC ATATAAATCT CCAGACCACC TTCCTTGGTG	3900
TACTTGAGAC TGTTTGAGAT GATTGCTCA ATAACCACTA GCAGCCACTT TTTATCCGTC	3960
ACGATTCTTT TATCAAGGTC ATGTAGATTG ACATTTAAGC CTTTTTGAAT AAAGAAAAGA	4020
GCATATTTAC GAATTATTTC CTTGACCAAG TCCTCAATTT GAACCTGCTT TAAGACCAAA	4080
TCATCATGGA AACTTTCTAA ACGCAGGTAC TGTAATACTA GGTGGGTATA GGAGTCGATT	4140
TTGAAAATTT CCTGTTCTAG CTGCTGCTTC AGTTGGCGGT CGACCACTTC TGCAACTAAG	4200
AGTTGACTGG CTGCAATGGG GGTCTTTATC TGATGGACCC ACAAGGTATA GTAATCCAGC	4260
AAATCCGTCG GTTTTCTTTC TGCTTTTGAC CTCTGCTGAT AGAGTTCCAT CTCACGCGCT	4320
TCTAATTTTT CTGCTAAAGC TATTTCCAAA GGAGACTTGG CTTCCCTCTC TCCATAGAGA	4380
AGTTCCTGGC GATAGACCTG CGTTTCCACC AATATGTCCC AAGTGAAAAA TAATATGGTT	4440
ACAAAGCAAC ACAAGAAGAA AAAGTAGAGG AAGTAAATTC CTAGACTGGC AAATAAAAAAC	4500
TGAAAGAGTA AGACAAGAAA TGCCAAAGAA AGCAGATAGA TAAAAAGACG ACTACGGGAG	4560
CGCAGATAGG CTAGAAAAAA TTGTTTCCAA TCAAGCATGC TTCAATCCGT ACCCTATTCC	4620
TTTCTTGGTC TCGATAAATC CTACCAATCC CTGCTCCTCC AACTTTTAC GCAAACGAGC	4680
CACATTGACA GAGAGGGTAT TATCATCAAT GAAAAAGTCA CTGTTCCAAA GTTCCCGCAT	4740
CAGGTCGTCA CGTGCTACGA TGTGCGCTGC ATGCTCAAAT AACACGCGTA AAATCTGGAA	4800
TTCATTTCTG GTCAAATTCA AGACTTGCCC TTGATAATGT AAATCCATGG ATTTGGTATT	4860
GAGGATAACA CCAGCATATT CCAGCAAACCT CTCATCACGC CCAAATCAT AGGAACGACG	4920
CAACAAGCCC TGAACCTTAG CTAAAAGAAC CTGCTGGTCA AAAGGCTTGG TCACAAAGTC	4980
ATCCGCCCCC ATATTGATTG CCATGACAAT ATCCATAGCC TGCTCTCTCG AAGAAAGAAA	5040
CATGATAGGT ACCTTGAAAA TCTTGCGGAT TTCCTGACAC CAGTGATAAC CATTAACAA	5100
GGGCAACCA ATATCCATGA GGACCAGATG AGGTTCCGAC TGAACAAATA GACTCAAAAC	5160

161

TTCCATAAAG TCTTCTACCA GGACCACTTC AAATCCCCAT TCAGAGAGCA TTTTCCCAAT	5220
CTGTTGACGA ATGACCTGAT CATCTTCTAT TAATAAAATC TTGTGCATGC GCTTCTCCTT	5280
TTCCATTATT ATAACAGATT TTTCCATGCT AGATGGTCTG AAACCTGAAT TGAAATAGCC	5340
TGTTTTTAGC CAGTACAAAC AGGCTATGCT ACTAGCTAAT TTGAGGGAAA TTTGCTAAGA	5400
TAAATAAAAA GAAAGGAGCT CTTATGGCCA ATATTTTGA CTATCTGAAA GATGTCGCAT	5460
ATGATTCTTA TTACGACCTT CCCTTGAATG AGTTAGACAT TCTAACCTTA ATAGAAATCA	5520
CCTACCTCTC CTTTGATAAT CTGGTCTCCA CACTTCCTCA ACGTCTTTTA GATCTAGCAC	5580
CTCAGGTTC AAGAGATCCC ACCATGCTTA CTAGCAAAAA TCGCCTTCAA TTATTAGATG	5640
AATTGGCTCA ACACAAGCGC TTCAAAAATT GCAAACTCTC CCATTTTATC AACGACATCG	5700
ACCCTGAACT GCAAAAGCAA TTTGCGGCTA TGACTTATCG TGTCAGCCTC GATACCTATC	5760
TGATTGTCTT TCGTGGGACA GATGACAGTA TCATTGGCTG GAAGGAAGAT TTCCACCTGA	5820
CCTATATGAA GGAAATTCCT GCTCAAAAGC ACGCCCTTCG CTATTTAAAG AACTTTTTTG	5880
CCCATCATCC TAAGCAAAAG GTTATTCTAG CTGGGCATTG CAAGGGAGGA AATCTCGCTA	5940
TCTATGCTGC TAGCCAAATT GAGCAAAGTT TGCAAAATCA GATCACAGCA GTTTATACAT	6000
TTGATGCACC TGGTCTCCAT CAAGAATTGA CACAGACTGC GGGTTATCAA AGGATAATGG	6060
ATAGAAGCAA GATATTCATT CCACAAGGTT CCATTATCGG TATGATGCTG GAAATTCCTG	6120
CTCACCAAAT CATCGTTCAG AGTACTGCCC TGGGTGGCAT CGCCCAGCAC GATACCTTTA	6180
GTTGGCAGAT TGAGGACAAG CACTTCGTCC AACTGGATAA GACCAACAGT GATAGCCAGC	6240
AAGTAGACAC AACCTTTAAA GAATGGGTGG CCACAGTCCC TGACGAAGAA CTTAGCTCT	6300
ACTTCGACCT CTTCTTTGGC ACTATTCTTG ATGCTGGTAT TAGCTCTATC AATGACTTGG	6360
CTTCCTTAAA GCGCTTGAA TACATTATC ATCTCTTTGT CCAAGCTCAA TCCCTCACTC	6420
CAGAAGAAAG AGAAACCTTG GGTGCGCTTA CCCAGTTATT GATTGATACT CGTTACCAGG	6480
CATGGAAAAA TAGATAATAC TCTTGAAAAT TAAATGTATA CAAAACAAAA GACCTAGAAT	6540
ACATACTTTC ATGTGCATT TAAGTCTTTT TAAATAGAAT CTAATAGTCA ATAAAAATCA	6600
AAGAGCATTG AGAGATAATG GGGCTTGAA CGTCCCTCTC GCTTCAACAA AATGACCCCA	6660
TTATAGATTA AAAAGATGCC ACTTAGAAAA AGCAAAAAAG GAAGTAAGAC AAAGGCAAAT	6720
ATATAAAAAG CTAAGTGAAC ATTCTCGTAT CCATTTTAT AAAAAAGGTA GGATAGATAA	6780
AAATAACTTG AAATGAGGGA TAATAAAAAT AATACTGGAT TCCACAAACT TCTATTATCC	6840
TTCCAAAATG ACACTATAAA GGCTAATACA ATTCCTATAA CGAGATACAT TTCTTACTCC	6900

162

TTTAATAGCT ACATTTTATC ATAATTATCC AAAGAAAAA GAGGGCATTT ATCCCTCTTA	6960
ATCCTTCATC TGA CTCTCTG CATCGGCCAC GACTTTTCT AGACTGGTTT GACCAAGTTC	7020
TGCCTCCATA GTCAACTGAA TTCTCTCCAA TTTTGTGATCC AAAACATCAT GAATATGAGC	7080
TCCTACAGGG CAATTGGAT TCGGATTGTC ATGGAAACTG AAGAGTTGAC CTGTCTTACC	7140
AAGACATTCG ACCGCCTGAT AAACATCTAA AAGACTAATA TCCTTAAGGT CCTTGACAAT	7200
CTCTGTTCCG CCCGTTCCAC GCGCTACTGA AATCAGCTCT GCCTTCTTCA ACTGGGACAA	7260
GATCTTTCTG ATAATGACAG GATTGACCCC GACACTAGCA GCCAGAAAAT CACTGGTCAC	7320
CTTGCTTTCC TTCCCTCGA GGGCAATGAT TATCAGCATA TGAGTCGCAA TGGTAAATCT	7380
ACTTGGAAAT TGCATCCTCT TCTCCTTTT ACGAGGCTAC CCTGCCTCTA CTCTTCTTTT	7440
TCTATTATTA TACCTTTT AGTTGTAATG TCAATCGTTA CCACTTTTCA ACCAGTCGTC	7500
TAACTCCCGA TCGCAGCCCT CTTTCTGAGC CAATTCTCTC AAAAATTCCT GATGATGAGT	7560
ATGGTGGATC CCATTGACCA GACTTTCATA GTAAACCTCA AAATAGGGAA GTCTCAGGTC	7620
TTTAGCCAGC TGCAATTCAG CTGCTACATC GTAGTCTACC CGTCGGAAGT CCATATCTAC	7680
CAGGCCTTTG TCATCAAACT CCAAAATCAT ATACTGGGCC CGCAAGTCCT TCCGTAGCTG	7740
AGCGTCCAAA AAGAAAGGTT GGCCAATCGA ACCCGGATTG ACAATCAATT GCCCACCAGT	7800
CCCGTAACGA AGCAACTGCT GGTGAATATG TCCATAAACA GCAATATCAC AGGGAGGATG	7860
AGTCACCAAG CGGTCAAACCT CCTCTGTGTT GCCAGTATGA ATCAACTCTC GCCCCAGTT	7920
CTTATCAGGC AGATGATGGC TAATTCCCAC CGTCAAATCC CCAAACTGAC GATGAATTTG	7980
AAGAGGTGA TTGTGGAGCA CTTCAATTC TTCTAGGGAA ATTTCTCTA AAACATACTG	8040
GCACTGGCGC AAGAGATAGC GTTGACTGGG GCGAGTACTG TCCAATTCCT TACGGACACC	8100
ATGCCAAAGA CTGTCTTCCC AGTTTCCCAA AACTCTAGCC GTAATCGGTA GTTGATCCAA	8160
CAAGTCCAAA ATCCTTCTAC GCCCTGTCCC TGGCATGAGA ATATCTCCCA AAAGCCAGTA	8220
TTTATCCACT CCTATCTGCC GAGCATCTGC CAAAACAGCC TCCAAGGCGG TGGTATTTCC	8280
ATGAATATCT GAAAGAAGAG CTATTTTCGT CATATCCATC TCCTCGTTT TTCTCTTGCA	8340
ATAAGTATAA CATAAAAAGT CACAGCTAGA GAAATCTAGC TTTTTTGTAT ATACTAGATA	8400
AAGATATTAG ACAAGAGGAA ACGAATGACC CCAAACAAAG AAGACTATCT AAAATGTATT	8460
TATGAAATG GCATAGACCT GCATAAGATT ACCAACAAGG AAATTGCGGC TCGCATGCAA	8520
GTCTCTCCCC CTGCCGTAAC TGAAATGATC AAACGAATGA AAAGTGAAAA TCTCATCCTA	8580
AAGGACAAGG AATGTGGCTA TCTACTGACT GACCTCGGTC TCAAACTGGT CTCTGAGCTC	8640
TATCGTAAGC ACCGCTTGAT TGAAGTTTT CTAGTTCATC ATTTAGACTA TACAAGTGAC	8700

163

CAGATTCACG AGGAAGCTGA GGTCTTGGAA CACACTGTCT CTGACCTGTT CGTGGAAAGA	8760
CTAGATAAAC TGCTAGGTTT CCCTAAAACC TGCCCCCAG GGGGAACTAT TCCTGCCAAG	8820
GGAGAACTAC TCGTTGAAAT CAATAACCTC CCACTAGCTG ATATCAAGGA AGCTGGCGCC	8880
TACCGCCTGA CTCGGGTGCA CGATAGTTTT GACATTCTCC ATTATCTGGA CAAGCACTCA	8940
CTTCACATCG GTGACCAGCT CCAAGTCAAG CAGTTTGATG GCTTCAGCAA TACCTTCACT	9000
ATCCTCAGTA ACGACGAGGA TTTACAAGTG AATATGGACA TTGCAAAACA ACTCTATGTC	9060
GAGAAATCA ACTAATTCT CAAGTCCCCT ACCAACCCTG AAAGTTTAT TTTGGCTCTT	9120
TGTCAACTGT AGTGGGTGA AGTCAGCTAA GCTCGAGAAA GGACAAATTT TGTCTTTCT	9180
TTTTTGATAT TCAGAGCGAT AAAATCCGT TTTTGAAGT TTTCAAAGTT CCGAAAACCA	9240
AAGGCATTGC GCTTGATAAG TTTGATGAGA TTATTGGTCG CTTCCAGTTT GGCATTAGAA	9300
TAGTGTAGTT GAAGGGCGTT GACAATCTTT TCTTTATCTT TGAGGAAGGT TTTAAAGACA	9360
GTCTGAAAAA TAGGATGAAC CTGCTTTAGA TTGTCTCAA TGAGTCCGAA AAATTTCTCC	9420
GGTTTCTTAT TCTGAAAGTG AAACAGCAAG AGTTGATAGA GCTGATAGTG GTGTTTCAAG	9480
TCTTGTGAAT AGCTCAAAAG CTTGTCTAAA ATCTCTTTAT TGGTTAAGTG CATACGAAAA	9540
GTAGGACGAT AAAATCGCTT ATCACTCAGT TTACGGCTAT CCTGTTGTAT GAGCTTCCAG	9600
TAGCGCTTGA TAGCCTTGTA TPCATGGGAT TTTCGATCCA ATTGGTTCAT AATTGAACA	9660
CGCACACGAC TCATAGCAGC GCTAAGATGT TGTACAATGT GAAAGCGATC CAACACGATT	9720
TTAGCATTCG GGAGTGAAAC AGTCTGGGAG ACTGTTTCAG CCTGAGCCTA GAAATTTGAA	9780
AGCGAAGCTG TTTAGCCAAG TCATAGTAAG GACTAAACAT ATCCATCGTA ATGATTTTCA	9840
CTTGACAACG AACGGCTCTA TCGTAGCGAA GAAAGTGATT TCGGATGACA GCTTGTGTTT	9900
TGCCTTCAAG AACAGTGATA ATATTAAGAT TATCAAAATC TTGCGCAATG AAACATCTCT	9960
TTCCCTTAGT GAAGGCATAC TCATCCCAAG ACATAATCTT TGGAAGCCGA GAAAAATCAT	10020
GCTCAAAGTG AAAGTCATTG AGCTTGCGAA TGACAGTTGA AGTTGAAATG GCCAGCTGAT	10080
GGGCAATATC AGTCATAGAA ATTTTTTCAA TTAACCTTTG AGCAATyTTT TGGTTGATGA	10140
TACGAGGGAT TTGGTGATTT TTCTTTACCA GGGGAGTCTC AGCAACCATC ATTTTTGAAC	10200
AGTGATAGCA CTTGAAACGA CGCTTTCTAA GGAGAATTCT AGAAGGCATA CCAGTCGTTT	10260
CAAGATAAGG AATTTTAGAA GGTTTTTGAA AGTCATATTT CTTCAATTGG TTTCCGCACT	10320
CAGGGCAAGA TGGGGCGTCG TAGTCCAGTT TGGCGATGAT TTCCTTGTGT GTATCCTTAT	10380
TGATGATGTC TAAAATCTGG ATATTAGGGT CTTTAATGTC TAGTAATTTT GTGATAAAAT	10440

164

GTAATTGTTC CATATGATTC TTTCTAATGA GTTGTTTGT CGCTTTTCAT TATAGGTCAT	10500
ATGGGACTTT TTTCTACAA TAAATAGGC TCCATAATAT CTATAGTGGA TTTACCCACT	10560
ACAAATATTA TAGAACCCTA AAAATAGAAG GAGATAGCAG GTTTCGAAGC CTGCTATCTT	10620
TTTTTGATGA CATTCAGGCT GATACGAAAT CATAAGAGGT CTGAACTAC TTTCAGAGTA	10680
GTCTGTTCTA TAAATATAG TAGATTGAAA TAAGATGTGA ACAACTCTAT CAGGAAAGTC	10740
AAATTAATTT ATAGAATTAT TTTAGCAGTC AAGGTGTACT GTTATAGATT CAATATATTA	10800
TATGACTATT AACCTTGCT TCTCCTAAAA TTGACTTTCT TGTTTTCTTA TCTTGTCAC	10860
TCGAAACAAG TATTGTAAGA ATTTGATTAT TTTTGAAAGT ACTTTTAATA TACTTGATAT	10920
AGTTAAAAAA GATTTGAAAC TAAATTCCAA ATTAGAAAAA GACTTGAAAT ACTAAAAAA	10980
AAAAAGTATA CTCTAATTGA AAACGGTAAC AAAACTAATT TAGAGAATGA AATATAGAGT	11040
ATTTCTCTCT TAAAAGTTT TGGTGAAACG AGATGTAGAA AGGAGATTTA GCCAAAGAGT	11100
CTATTAGTGC TAGAATAATA GATTAGAATT ATTTTAGAAA AACGAAGTGA GCAGCTTATA	11160
AATTCGAAGC CCCAAATAGA TTCATACTAG TATCTTTGCG AAAAAATAAA GGGCGACTTC	11220
CTTCATGAAT ATCAATTTCA TCTATAAGGA AGGTAGCTAA TTGAACTAAC TTATTTATTC	11280
TGTTTGTGCG TAGAAAAATC AGACCTCCTT GTGAAGATTG AGGAGATACT TAATGAAAT	11340
CAAAGAAGAA ACTAGCAAGC TAGTAGCAGA TTGCCCAAAA CACCGCTTTG AGGTTGTAGA	11400
TAAGACTGAC CTATATAATC CAAGGTGAAG CGACTGTGGT TTGAAGAGAT TTTCAAAGAG	11460
TATAGGCTAG AGAGTAGTGT TTTTATGTCC TTCTAGTAGA AAATGCTAGA CAGAAGAATG	11520
GGGAACCTGG ATAGGAAAAA TAGATTGAGA AAGGAGGTTA GAAGAGATGA TTATTACAAA	11580
AATTAGCCGT TTAGGAACCT ATGTGGGAGT AAATCCACAT TTTGCAACAT TAATAGATTT	11640
TCTAGAAAAA ACAGGACTAG AAAATTTAAC AGAAGGTTGCT ATTGCTATCG ATGGTAATCG	11700
ATTGTTTGGG AATTGCTTTA CTTATCTAGC AGATGGTCAA GCAGGGGCTT TCTTTGAAAC	11760
CCACCAAAAA TATTGGGATA TTCATTTAGT TTTGGAAAAA GAAGAAGCCA TGGCTGTTAC	11820
ATCGCCGGAA AATGTAAGCG TTACCCAAGA ATATGATGAA GAGAAAGATA TTGAATTATA	11880
CACAGGGAAA GTGGAACAGT TGGTTCATTT GAGAGCTGGC GAATGCCTCA TCACTTTTC	11940
AGAAGATTTA CATCAACCCA AGGTTCTGAT AAATGATGAA CCTGTGAAAA AAGTTGCTTT	12000
TAAAGTTGCG ATTTCTTAAT GTAGAAAGAG AAGAACGATG AAAAAAATGA GAAAGTTT	12060
ATGTCTAGCT GGAATGCGC TAGCGGCTGT TGCCTTGGTA GCTTGTTCAG GAAAAAAGA	12120
AGCTACAAC AGTACTGAAC CACCAACAGA ATTATCTGGT GAGATTACAA TGTGGCACTC	12180
CTTTACTCAA GGACCCCGTT TAGAAAGTAT TCAAAAATCA GCAGATGCTT TCATGCAAAA	12240

165

GCATCCAAAA	ACGAAAATCA	AGATTGAAAC	ATTTCTTGG	AATGACTTCT	ATACTAAATG	12300
GACTACAGGT	TTAGCAAATG	GAAATGTGCC	AGATATCAGT	ACAGCTCTTC	CTAACCAAGT	12360
AATGGAAAATG	GTCAACTCAG	ATGCTTTGGT	TCCGCTAAAT	GATTCTATCA	AGCGTATTGG	12420
ACAAGATAAA	TTTAACGAAA	CTGCCTTAAA	TGAAGCAAAA	ATCGGAGATG	ATTACTACTC	12480
TGTTCTCTTT	TATTCACATG	CACAAGTCAT	GTGGGTTAGA	ACAGATTTGT	TAAAAGAACA	12540
TAATATMGAG	GTTCCATAAA	CTTGGGATCA	ACTCTATGAA	GCTTCTAAAA	AATTGAAAGA	12600
AGCTGGAGTT	TATGGCTTGT	CTGTTCCGTT	TGGAACAAAT	GACTTAATGG	CAACACGTTT	12660
CTTGAAGTTC	TACGTACGTA	CTGGTGGAGG	AAGCCTCTTA	ACAAAAGATC	TTAAAGCAGA	12720
CTTGACAAGC	CAACTTGCTC	AAGATGGTAT	TAAATACTGG	GTTAAATTGT	ATAAAGAAAT	12780
CTCACCTCAA	GATTCTTTGA	ACTTTAATGT	CCTTCAACAA	GCTACCTTGT	TCTATCAAGG	12840
AAAAACAGCA	TTTGACTTTA	ACTCTGGCTT	CCATATCGGA	GGAATTAATG	CCAACAGTCC	12900
TCAATTGATT	GATTCGATTG	ATGCTTATCC	TATTCCAAAA	ATCAAAGAGT	CTGATAAAGA	12960
CCAAGGAATT	GAAACCTCAA	ACATTCCAAT	GGTTGTTTGG	AAAAATTCAA	AACATCCAGA	13020
AGTTGCTAAA	GCATTCTTAG	AAGCACTTTA	TAATGAAGAA	GACTACGTTA	AATTCCTTGA	13080
TTCAACTCCA	GTAGGTATGT	TGCCAACTAT	TAAGGGGATT	AGCGATTCTG	CAGCCTATAA	13140
AGAAAAATGAA	ACTCGTAAGA	AATTTAAACA	TGCTGAAGAA	GTAATTACTG	AAGCTGTAA	13200
AAAAGGTACT	GCTATTGGTT	ATGAAAAATG	GCCAAGTGTA	CAAGCTGGTA	TGTTGACTAA	13260
CCAACACATT	ATTGAACAAA	TGTTCCAAGA	TATCATTACA	AATGGAACAG	ATCCTATGAA	13320
AGCAGCAAAA	GAAGCAGAAA	AACAATTAAA	TGATTTATTT	GAGGCTGTTC	AGTAGATGTA	13380
AAAGACTAGA	AAATAGGTGG	GATAGTGAGC	TGAAAAGCTC	TAGCCCAATC	TTGTAAAAGA	13440
AGGGAGAAGG	AGAATGGTTA	AAGAACGTAA	TTAACTCGC	TGGATATTTG	TTTTGCCAGC	13500
TATGATTATC	GTAGGATTAC	TCTTTGTTTA	TCCGTTTTC	TCGAGTATTT	TTTATAGCTT	13560
TACCAATAAG	CATTTGATTA	TGCCTAATTA	TAAATTTGTT	GGTTTGGCTA	ACTATAAAGC	13620
TGTGCTATCA	GATCCCAACT	TCTTTAATGC	GTTCTTTAAT	TCAATTAAAT	GGACCGTTTT	13680
CTCATTAGTT	GGTCAAGTTT	TAGTAGGGTT	TGTATTGGCT	TTAGCTCTTC	ACAGAGTACG	13740
CCACTTCAAG	AAATTATATA	GGACATTATT	GATTGTTTCT	TGGGCATTTT	CTACCATCGT	13800
TATTGCCTTC	TCTTGGCAGT	GGATTCTAAA	CGGGGTTTAT	GGCTACTTAC	CTAATCTAAT	13860
CGTAAAATTA	GGTTTAATGG	AACATACACC	TGCATTTTTC	ACAGATAGTA	CATGGGCATT	13920
CCTATGTTTG	GTGTTTATCA	ACATTTGGTT	TGGAGCACCA	ATGATTATGG	TTAATGTGCT	13980

166

TTCAGCTTTG CAAACAGTAC CAGAAGAACA ATTTGAGGCT GCTAAGATAG ATGGTGCTTC	14040
AAGTTGGCAG GTGTTCAAGT TTATCGTCTT TCCACATATT AAAGTGGTTG TAGGACTTCT	14100
AGTTGTTTTG AGAACTGTAT GGATCTTTAA TAACTTTGAC ATTATCTACC TCATTACTGG	14160
TGGTGGACCA GCCAATGCTA CAACGACGCT TCCAATTTTT GCTTACAACC TGGGCTGGGG	14220
AACTAAATTG TTGGGTCGTG CTTCAGCAGT TACAGTACTG CTCTTTATCT TCTTGGTGGC	14280
GATTTGCTTT ATCTACTTTG CTATCATCAG TAAGTGGGAA AAGGAGGGTA GAAAATAATG	14340
AAGAAGAAAT CCAGTATTTA TTTAGATATT CTCTCACATG TACTTTTACT TGGTGCGACC	14400
ATCGTTGCAG TTTTCCCATT GGTATGGATT ATCATATCTT CTGTCAAAGG GAAAGGGGAA	14460
TTAACTCAGT ATCCAACACG ATTTTGGCCT GAACAGTTTA CATTAGATTA TTTCACTCAT	14520
GTTATCAACG ATTTGCACTT CATTGATAAC ATTCGAAACA GTTTAATCAT TGCCTTGGCT	14580
ACAACCCTTA TTGCGATTAT TATTTCTGCT ATGGCAGCCT ATGGTATTGT TCGATTCTTT	14640
CCTAAATTGG GAGCAATCAT GTCGAGACTA CTCGTCAATTA CCTACATTTT CCCACCAATT	14700
TTGTTAGCAA TTCCCTATTC AATTGCCATT GCTAAAGTTG GGTTAACAAA TAGTTTATTT	14760
GGCTTGATGA TGGTTTATCT ATCTTTTAGT GTTCCATATG CAGTTTGGCT CTTAGTTGGA	14820
TTTTTCCAAA CAGTTCCAAT TGGAAATTGAA GAAGCGGCTA GAATTGATGG TGCAAAATAA	14880
TTTGTACGT TTTATAAAGT TGTGCTACCG ATTGTAGCAC CAGGTATTGT AGCAACAGCT	14940
ATTTATACAT TTATCAATGC TTGGAATGAA TTCCTGTATG CCTTGATTTT GATTAACAAT	15000
ACAGGAAAGA TGACAGTAGC AGTAGCCCTT CGTTCACCTA ATGGTTCAGA AATACTAGAC	15060
TGGGGAGATA TGATGGCAGC GTCTGTTATT GTAGTTCTTC CATCAATTAT TTTCTTCTCT	15120
ATCATCCAAA ATAAGATTGC AAGTGGATTA TCAGAAGGAT CTGTGAAGTA GACGAAAGAA	15180
GGAAAAAAT GAATAAAAGA GGTCTTTATT CAAAAGTAGG AATTTCCGTT GTAGGCATTA	15240
GTCTTTTAAT GGGAGTCCCC ACTTTGATTC ATGCCAATGA ATTAACTAT GGTCAACTGT	15300
CCATATCTCC TATTTTTCAA GGAGGTTTAT ATCAACTGAA CAATAAGAGT ATAGATATCA	15360
GCTCTTTGTT ATTAGATAAA TTGTCTGGAG AGAGTCAGAC AGTAGTAATG AAATTTAAAG	15420
CAGATAAACC AAACCTCTCT CAAGCTTTGT TTGGCCTATC TAATAGTAAA GCAGGCTTTA	15480
AAAATAATTA CTTTTCATTT TTCATGAGAG ATTCTGGTGA GATAGGTGTA GAAATAAGAG	15540
ACGCCCCAAA GGAATAAAT TATTTATTTT CCAGACCAGC TTCATTATGG GGAACATA	15600
AAGGACAGGC AGTTGAAAAT ACACTAGTAT TTGTATCTGA TTCTAAAGAT AAAACATACA	15660
CAATGTATGT TAATGGAATA GAAGTGTCTT CTGAAACAGT TGATACATTT TTGCCAATTT	15720
CAAATATAAA TGGTATAGAT AAGGCAACAC TAGGAGCTGT TAATCGTGAA GGTAAGGAAC	15780

167

ATTACCTCGC AAAAGGAAGT ATTGATGAAA TCAGTCTATT TAACAAAGCA ATTAGTGATC	15840
AGGAAGTTTC AACTATTCCC TTGTCAAATC CATTTTCAGTT AATTTTCCAA TCAGGAGATT	15900
CTACTCAAGC TAACTATTTT AGAATACCGA CACTATATAC ATTAAGTAGT GGAAGAGTTC	15960
TATCAAGTAT TGATGCACGT TATGGTGGGA CTCATGATTC TAAAAGTAAG ATTAATATTG	16020
CCACTTCTTA TAGTGATGAT AATGGGAAAA CGTGGAGTGA GCCAATTTTT GCTATGAAGT	16080
TTAATGACTA TGAGGAGCAG TTAGTTTACT GGCCACGAGA TAATAAATTA AAGAATAGTC	16140
AAATTAGTGG AAGTGCTTCA TTCATAGATT CATCCATTGT TGAAGATAAA AAATCTGGGA	16200
AAACGATATT ACTAGCTGAT GTTATGCCTG CGGTATTGG AAATAATAAT GCAAATAAAG	16260
CCGACTCAGG TTTTAAAGAA ATAAATGGTC ATTATTATTT AAAACTAAAG AAGAATGGAG	16320
ATAACGATTT CCGTTATACA GTTAGAGAAA ATGGTGTCGT TTATAATGAA ACAAATAATA	16380
AACCTACAAA TTATACTATA AATGATAAGT ATGAAGTTTT GGAGGGAGGA AAGTCTTTAA	16440
CAGTCGAACA ATATTCGGTT GATTTTGATA GTGGCTCTTT AAGAGAAAGG CATAATGGAA	16500
AACAGGTTCC TATGAATGTT TTCTACAAAG ATTCTGTATT TAAAGTGACT CCTACTAATT	16560
ATATAGCAAT GACAACTACT CAGAATAGAG GAGAGAGTTG GGAACAATTT AAGTTGTTGC	16620
CTCCGTCTTT AGGAGAAAAA CATAATGGAA CTTACTTATG TCCCGGACAA GGTTTAGCAT	16680
TAAATCAAG TAACAGATTG ATTTTGTCAA CATATACTAG TGGAGAACTA ACCTATCTCA	16740
TTTCTGATGA TAGTGGTCAA ACATGGAAGA AATCCTCAGC TTCAATTCCG TTTAAAAATG	16800
CAACAGCAGA AGCACAAATG GTTGAAGTGA GAGATGGTGT GATTAGAACA TTCTTTAGAA	16860
CCACTACAGG TAAGATAGCT TATATGACTA GTAGAGATTC TGGAGAAACA TGGTCGAAAG	16920
TTTCGTATAT TGATGGAATC CAACAAACTT CATATGGCAC ACAAGTATCT GCAATTAAAT	16980
ACTCTCAATT AATTGATGGA AAAGAAGCAG TCATTTTGAG TACACCAAAT TCTAGAAGTG	17040
GCCGCAAGGG AGGCCAATTA GTTGTGCGTT TAGTCAATAA AGAAGATGAT AGTATTGATT	17100
GGAAATACCA CTATGATATT GATTGCTT CGTATGGTTA TGCCTATTCT GCGATTACAG	17160
AATTGCCAAA TCATCACATA GGTGTACTGT TTGAAAAATA TGATTCGTGG TCGAGAAATG	17220
AATTGCATTT AAGCAATGTA GTTCAGTATA TAGATTTGGA AATTAATGAT TTAACAAAAT	17280
AAAGGAGAAA AACATGGTTA AATACGGTGT TGTGGAACA GGGTATTTTG GAGCTGAATT	17340
GGCTCGCTAC ATGCAAAAGA ATGATGGAGC AGAGATTACT CTTCTCTATG ATCCAGATAA	17400
TGCAGAGGCG ATTGCAGAAG AATTGGGAGC AAAAGTAGCA AGTTCCTTAG ATGAGTTGGT	17460
TTCTAGCGAT GAAGTAGATT GTGTTATCGT CGCAACTCCA AATAATCTTC ATAAGGAACC	17520

168

GGTTATTAAG GCTGCACAGC ATGGTAAAAA TGTTTTCTGT GAAAAACCAA TTGCGCTTTC	17580
TTATCAAGAT TGTCGCGAGA TGGTAGATGC GTGTAAAGAA AACAAATGTAA CCTTTATGGC	17640
AGGACATATT ATGAATTTCT TTAATGGTGT TCATCATGCA AAAGAACTCA TTAATCAAGG	17700
AGTTATCGGA GACGTTCTAT ATTGTCATAC AGCTCGTAAT GGTGGGAAG AACAAACACC	17760
GTCAGTATCA TGGAAAAAAA TTCGTGAAAA ATCAGGTGGT CACTTGATC ACCACATCCA	17820
TGAATTGGAT TGCCTTCAAT TCCTTATGGG GGGCATGCCT GAAACTGTAA CCATGACAGG	17880
TGGAAATGTG GCCCATGAAG GTGAACATTT CGGTGATGAA GATGATATGA TTTTGTCAA	17940
TATGGAATTT TCTAATAAGC GTTTGCCTT GTTAGAATGG GGTTCAGCTT ATCGTTGGGG	18000
TGAACATTAT GTCTTAATCC AAGGAAGCAA AGGTGCCATC CGCTTAGACT TATTCAACTG	18060
TAAAGGAACT CTTAAGCTAG ATGGGCAAGA AAGCTATTTT TTGATTCACG AATCGCAAGA	18120
AGAAGATGAT GATCGGACTC GTATCTATCA TAGTACAGAG ATGGATGGAG CAATTGCTTA	18180
TGGTAAACCA GGTAAACGTA CTCCATTATG GCTATCATCT GTCATTGATA AAGAAATGCG	18240
CTATCTGCAT GAGATTATGG AAGGAGCTCC AGTATCAGAA GAATTTGCAA AACTTTTGAC	18300
AGGTGAAGCT GCCCTAGAAG CAATTGCTAC TGCAGATGCT TGTACCCAGT CTATGTTTGA	18360
AGATCGCAAA GTAAATTTGT CAGAAATGT AAAATAAATT TTGGTATTCT CCTATTATA	18420
GGTCGACTTG CTCCTCTGAA AGTACTTTTA GAGGAGCTGT TTGACTTTGC TAGTTTTGA	18480
AACTGAAATC TATTATACTA CAAACTATTG AAAGCGTTTT AATTTTAAGG TATAATAATC	18540
TCATAGAAAT AAAGAAAAGG AGGAAAGAGG ATGCCACAGA TTAGCAAAGA AGCCTTGATT	18600
GAGCAAATCA AAGATGGAAT CATCGTTTCT TGTCAGGCTC TTCCTCATGA ACCGCTTAT	18660
ACAGAAGCGG GAGGGGTGAT TCCCTTGCTG GTCAAAGCGG CTGAGCAAGG TGGAGCAGTC	18720
GGTATCCGAG CAAACAGTGT TCGCGATATC AAGGAAATTA AGGAAGTCAC TAAACTTCCA	18780
ATCATGCGGA TTATCAAACG TGATTATCCA CCTCAGGAAC CCTTCATCAC GGCTACTATG	18840
AAAGAAAGTTG ATGAATTGGC AGAACTGGAC ATCGAGGTGA TTGCTCTGGA TTGTACCAAG	18900
CGTGAACGCT ACGATGGTTT GGAAATTCAA GAGTTCATTC GTCAGGTAA GGAGAAATAT	18960
CCTAATCAGC TTTTGATGGC TGATACTAGT ATCTTCGAAG AAGGGCTAGC AGCTGTAGAA	19020
GCAGGAATTG ACTTTGTCGG AACAACTTA TCAGGCTACA CATCCTACAG TCCAAAAGTA	19080
GACGGTCCAG ATTTTGAATT GATTAAGAAA CTCTGTGATG CTGGTGTAGA TGTCATTGCA	19140
GAAGGAAAAA TTCATACACC AGAACAAGCC AAACAAATCC TTGAATATGG AGTGCGAGGC	19200
ATCGTTGTTG GTGGCGCCAT TACTAGACCA AAAGAGATTA CAGAACGCTT CGTTGCTAGT	19260
CTTAAATAAG ATGTGAGGGG GAGTTTTATG TTTAAAGTTT TACAAAAAGT TGGAAAAGCT	19320

TTTATGTTAC CTATAGCTAT ACTTCCTGCA GCAGGTCTAC TTTTGGGGAT TGGTGGTGCA	19380
CTTTCAAACC CAACCACGAT AGCAACTTAT CCAATACTAG ACAATAGTAT TTTTCAATCA	19440
ATATTCCAAG TAATGAGCTC TGCAGGAGAG GTTGTATTCA GTAATTTGTC ACTACTTCTC	19500
TGTGTGGGAT TATGTATTGG CTTAGCGAAA CGAGATAAAG GAACCGCTGC GTTAGCAGGA	19560
GTAAGTGGTT ACTTAGTTAT GACTGCAACG ATCAAAGCTT TGGTAAAACT TTTTATGGCA	19620
GAAGGATCTG CAATTGATAC TGGAGTTATT GGAGCATTAG TTGTCGGAAT AGTTGCCGTA	19680
TATTGTCACA ACCGATATAA CAATATTCAA TTACCTTCCG CTTTAGGATT CTTTGGAGGT	19740
TCACGCTTCG TTCCTATTGT TACATCGTTC TCTTCTATCT TGATTGGCTT TGTCTTCTTT	19800
GTTATTTGGC CACCTTTCCA ACAACTTCTT GTTCTACAG GTGGATATAT TTCTCAGGCG	19860
GGTCCAATTG GAACTTTTCT ATATGGATT TTAATGAGAC TTTCTGGAGC AGTAGGCTTA	19920
CATCATATAA TTTACCCTAT GTTTTGGTAT ACTGAACTTG GTGGTGTGA AACTGTTGCA	19980
GGACAAACAG TGGTTGGAGC TCAAAAATA TTTTGTGCTC AATTAGCCGA TTTGCCCCAT	20040
TCTGGATTAT TTACAGAAGG AACAAAGTTT TTTGCAGGTC GTTCTCAAC AATGATGTTT	20100
GGTTTACCGG CTGCCTGTTT AGCGATGTAC CATAGTGTTC CTAAAAATCG TCGTAAAAAA	20160
TACGCGGGTT TGTTTTGTGG AGTTGCTTTA ACATCTTTTA TTACCGGTAT TACAGAACCA	20220
ATTGAATTTA TGTTCCTATT CGTCAGTCCG GTTCTATATG TTGTTACGCG ATTCCTTGAT	20280
GGTGTAGCT TCTTTATTGC AGACGTCTTA AATATTTCAA TAGGAAACAC ATTTTCAGGA	20340
GGTGTAAATCG ATTTCACTTT ATTTGGAATT TTGCAGGGGA ACGCTAAGAC GAATGGGTT	20400
CTTCAGATTC CATTTGGACT TATTTGGAGT GTTTTGTATT ATATTATTTT TAGATGGTTC	20460
ATTACTCAAT TCAACGTTCT AACGCCAGGG CGAGGAGAAG AAGTAGATTC TAAAGAAATT	20520
TCTGAATCCG CAGATTCAAC TTCAAATACT GCAGATTATT TAAAACAGGA TAGCCTACAA	20580
ATTATCAGAG CCTTGGGTGG ATCAAATAAT ATAGAAGATG TAGATGCTTG TGTGACACGT	20640
TTACGTGTAG CTGTAAAAGA AGTTAATCAA GTTGATAAAG CACTTTTAAA ACAAATTGGT	20700
GCAGTTGATG TCTTAGAAGT GAAGGGTGGC ATTCAAGCAA TCTATGGAGC AAAAGCAATC	20760
TTATATAAAA ATAGTATTAA TGAAATTTTA GGTGTAGATG ATTAAGTACT TACTGACTTA	20820
ATAAAAAACA GAGGAGAGTG ATGGATGAGT AGGATGAAAT GAAATCGCAT ACAAGAAATA	20880
AAGAAGTCAT TATCCAAGTT GGATACGCTT ATTACATAGG AGAATACAAA TGAAATTTAG	20940
AAAATTAGCT TGTACAGTAC TTGCGGGTGC TGCGGTTCTT GGTCTTGCTG CTTGTGGCAA	21000
TTCTGGCGGA AGTAAAGATG CTGCCAAATC AGGTGGTGAC GGTGCCAAAA CAGAAATCAC	21060

170

TTGGTGGGCA TTCCAGTAT TTACCCAAGA AAAAAGTGGT GACGGTGTG GAACTTATGA	21120
AAAATCAATC ATCGAAGCGT TTGAAAAAGC AAACCCAGAT ATAAAAGTGA AATTGGAAAC	21180
CATCGACTTC AAGTCAGGTC CTGAAAAAAT CACAACAGCC ATCGAAGCAG GAACAGCTCC	21240
AGACGTACTC TTTGATGCAC CAGGACGTAT CATCCAATAC GGTA AAAACG GTAAATTGGC	21300
TGAGTTGAAT GACCTCTTCA CAGATGAATT TGTTAAAGAT GTCAACAATG AAAACATCGT	21360
ACAAGCAAGT AAAGCTGGAG ACAAGGCTTA TATGTATCCG ATTAGTTCTG CCCCATTTCTA	21420
CATGGCAATG AACAAGAAAA TGTTAGAAGA TGCTGGAGTA GCAAACCTTG TAAAAGAAGG	21480
TTGGACAACT GATGATTTTG AAAAAGTATT GAAAGCACTT AAAGACAAGG GTTACACACC	21540
AGGTTCAATG TTCAGTTCTG GTCAAGGGG AGACCAAGGA ACACGTGCCT TTATCTCTAA	21600
CCTTTATAGC GGTCTGTAA CAGATGAAAA AGTTAGCAAA TATACAACCTG ATGATCTTAA	21660
ATTCGTCAAA GGTCTTGAAG AAGCAACTAG CTGGATTAAA GACAATTTGA TCAATAATGG	21720
TTCACAATTT GACGGTGGG CAGATATCCA AAACCTTGCC AACGGTCAAA CATCTTACAC	21780
AATCCTTTGG GCACCAGCTC AAAATGGTAT CCAAGCTAAA CTTTTAGAAG CAAGTAAGGT	21840
AGAAGTGGA GAAGTACCAT TCCCATCAGA CGAAGGTAAG CCAGCTCTG AGTACCTTGT	21900
AAACGGGTTT GCAGTATTC ACAAATAAGA CGACAAGAAA GTCGCTGCAT CTAAGAAATT	21960
CATCCAGTTT ATCGCAGATG ACAAGGAGTG GGGACCTAAA GACGTAGTTC GTACAGGTGC	22020
TTTCCAGTC CGTACTTCAT TTGGAAAAC TTATGAAGAC AAACGCATGG AAACAATCAG	22080
CGGCTGGACT CAATACTACT CACCATACTA CAACACTATT GATGGATTG CTGAAATGAG	22140
AACACTTTGG TTCCCAATGT TGCAATCTGT ATCAAATGGT GACGAAAAAC CAGCAGATGC	22200
TTTGAAAGCC TTTACTGAAA AAGCGAACGA AACAATCAAA AAAGCTATGA AACAATAGTC	22260
CTTAGTTATT CTATAAAAAG TAGTTTTTTA AAGAACCTAA GAGTGATATC CCCCTTTTCC	22320
CTCTACACAG ATAGTGTAAG AAAAGGGGGC TTTGTMTTAA AATGTAAGAA ACTGTACCGA	22380
AATTAAAATG AAGTTCCTTAC ATAAGCGAAT CATAAAAAAT TTCATTTTGA TTTTAAACA	22440
GTTCAGAAA GTCAAAAAAT TATTCTATTT GAAAGAGAGG TGCCGACTGT GAAAGTCAAT	22500
AAAATCCGTA TGCGGGAAAC AGTGATTTCC TACGCTTTCC TAGCACCAGT ATTATTCTTC	22560
TTTGTCTATCT TTGTGTGGC TCCGATGGT ATGGGCTTCA TTACAAGTTT CTTTAACTAC	22620
TCAATGACTA AATTTGAGTT TGTAGGCTTG GATAACTATA TCCGTATGTT TAAAGATCCT	22680
GTCTTTACAA AATCTCTGAT TAACACAGTT ATTTTGGTTA TTGGATCTGT ACCAGTTGTT	22740
GTCTATTCT CACTCTTGT AGCATCTCAG ACCTATCATC AAAATGTCAT TGCCAGATCC	22800
TTCTACCGT TCGTCTCTT CCTTCCTGTT GTAACGGTA GTGTTGCCGT GACAGTTGTT	22860

TGGAAATGGA TTTATGACCC ACTATCAGGG ATTCTAAACT TTGTCCTTAA GTCCAGCCAC	22920
ATCATCAGCC AAAACATTTT TTGGTTGGGA GATAAAAACT GGGCATTGAT GGCGATTATG	22980
ATTATTTCTCT TGACCACTTC AGTTGGTCAG CCCATCATCC TTTATATCGC TGCCATGGGG	23040
AATATTGACA ATTCAGTGGT TGAAGCGGCG CGTGTGTATG GTGCAACTGA GTTTCAAGTT	23100
TTTTGGAAGA TTAAATGGCC AAGCCTTCTT CCAACAACTC TTTATATTGC AATCATCACA	23160
ACAATTAAC CATTCCAGTG TTTCGCCTTG ATTCAGCTTT TGACATCTGG TGGTCCAAAC	23220
TACTCAACAA GTACCTTGAT GTACTACCTT TACGAAAAAG CCTTCCAATT GACAGAATAC	23280
GGCTATGCCA ACACAATTGG TGTCTTCTTG GCAGTCATGA TTGCTATCGT AAGCTTTGTT	23340
CAATTTAAAG TACTTGAAA CGACGTAGAA TACTAAAGAA AGGAGACAGC TATGCAATCT	23400
ACAGAAAAA AACCATTAAAC AGCCTTTACT GTTATTTCAA CAATCATTTT GCTCTGTGTG	23460
ACTGTGCTGT TCATCTTTCC ATTCTACTGG ATTTTGACAG GGGCATTCAA ATCACAACCT	23520
GATACAATTG TTATTCCTCC TCAGTGCTTC CCTAAAATGC CAACCATGGA AAACCTCCAA	23580
CAATCATGG TGCAGAACCC TGCCTTGCAA TGGATGTGGA ACTCAGTATT TATCTCATTG	23640
GTAACCATGT TCTTAGTTTG TGCAACCTCA TCTCTAGCAG GTTATGTATT GGCTAAAAA	23700
CGTTTCTATG GTCAACGCAT TCTATTTGCT ATCTTTATCG CTGCTATGGC GCTTCCAAA	23760
CAAGTTGTCC TTGTACCATT GGTACGTATC GTCAACTTCA TGGGAATCCA TGATACTCTC	23820
TGGGCAGTTA TCTTGCCTTT GATTGGATGG CCATTCGGTG TCTTCCTCAT GAAACAGTTC	23880
AGTGAAAATA TCCCTACAGA GTTGCTTGAA TCAGCTAAAA TCGACGGTTG TGGTGAGATT	23940
CGTACCTTCT GGAGTGTAGC CTTCCCGATT GTGAAACCAG GGTITGCAGC CCTTGCAATC	24000
TTACCTTCA TCAATACTTG GAATGACTAC TTCATGCAAT TGGTAATGTT GACTTCACGT	24060
AACAATTTGA CCATCTCACT TGGGGTTGCG ACCATGCAGG CTGAAATGGC AACCAACTAT	24120
GGTTTGATTA TGGCAGGAGC TGCCCTTGCT GCTGTTCCAA TCGTCACAGT CTTCTAGTC	24180
TTCCAAAAAT CCTTCACACA GGGTATTACT ATGGGAGCGG TCAAAGGATA ATACTCTGCG	24240
AAAATCTCTT CAAACTACGT CAGCTTCACC TTGCCATACT TAAGTATTGC CTGCGGTTAG	24300
CTTCCTAGTT TGTTCTTCAA TTTTCATTGA GTATAGGAAA ATCAATCTAT CAAGATACAG	24360
AAGTATATTT TATAGATTTA GAGAATATAG AGGTTATAAG TGTCTACAAA ATGGAGGGTA	24420
TGCAGTTACT TTATGAAGTT TTGTCAGACA CTTATAAACT TAAGAATGGT TTTAGTTAAC	24480
TATCAGAAAC GAAGGAAAGA GTATGATTTT TGACGATTTG AAAACATCA CCTTTTACAA	24540
AGGGATTCAT CCTAATTTAG ACAAGGCTAT CGACTATCTC TACCAACATC GTAAGGATTC	24600

172

TTTCGAATTA GGAAAGTATG ATATTGATGG AGATAAAGTC TTTCTAGTTG TTCAGGAAAA	24660
TGTCCTCAAT CAAGCTGAAA ATGATCAATT TGAGTATCAT AAGAACTATG CAGATTGCA	24720
TTTGCTGGTA GAAGGACATG AATATTCGAG CTACGGTTCA CGTATCAAAG ACGAGGCAGT	24780
AGCATTCGAC GAAGCGAGTG ACATTGGCTT TGTTCAATTGT CATGAACACT ACCCACTCTT	24840
GTTGGGTTAT CACAATTTTG CGATTTTCTT CCCAGGTGAG CCACATCAGC CAAATGGTTA	24900
TGCAGGCATG GAAGAAAAGG TTCGAAAATA TCTCTTTAAA ATTTTGATTG ATTA AAAATA	24960
GGATGAATTG TTTTGTGTA AAGCTTTGAT AATACTCTAC CATGAAATTG ATCTTTGTGA	25020
GGTAGAGAAA TGAGAATAAA ATATTTAAAA ATTGGTATCT TCTAAGTATG CTGCAAGAGC	25080
TAGTTTCTTA GATGGACAGG GGATTACAGT TGATGAGATG GCTTGGATAA TTAGGGGCAT	25140
TGTGAATGCA TTGATTGGTA GATACATAAA ATTAGGTACT TATGCGGCTA AGTATGGTAT	25200
TAGTATGGCA CGCTCGATCT TAAGTAGGGT AGCTGCAACT GCAGCAGCAA GAGTAGGATT	25260
ACTGACCAAG ATTTCTGGAT GGATTTTACG AGTAGCTGTG AATGTAGCTG ATGTATATGG	25320
TAATTTTGCC AACAAATATTG CTGCAGCTTG GGATGCATAT GATAAAATTC CTAACAATGG	25380
TCGTATAAAC TTTTAAATG CGAGAATGAA AGCACTTTGT ATTTTTTTAT TGAATATGTT	25440
AGCTTGGACA GTGCTTGCAA TGATAATTCTG TGGAGGGCTA GATGGATTG ATAGGCATAC	25500
TGGAGTACT ATTTTAATTG CGTCGCTGTT CGGGGTATAT GATTATAAGC CCATAGATAA	25560
AAATAGAAAA AAGTCCAAAA GAAAAAATAG ATTTGTTTAT GGTAGGGACT TATGAAAGCT	25620
TTACTGACAA AAAAGAAAAC AGTTTACAAA GAAAAATGAT GGAGGAGCAA ACATGGCACA	25680
AAAAGGAGTA AGCCTTATCA AGGCAGCATT TGATACAGAT AACTTTCTCA TGCCTTTTAG	25740
TGAGAAGGTC TTGGACATCG TGACAGCCAA TCTTCTTTTT GTCGTCTCTT GTTTACCCAT	25800
CGTGACGATT GGAGTGGCTA AAATCAGCCT CTACGAGACC ATGTTTGAAG TTAAGAAGAG	25860
CAGACGGGTG CCTGTTTTTA AAATCTATCT AAGATCTTTC AAGCAAAATC TGAAACTAGG	25920
TCTTCAGCTG GGTTTAATGG AGTTAGGAAT TGTGTTTCTT ACCCTTTCAG ATCTCTATCT	25980
TTTCTGGGGT CAAACAGCTC TGCCCTTCCA ATTGCTGAAA GCCATTTGTT TAGGTATTCT	26040
GATTTTCTT ACTATCGTGA TGCTGGCTAG TTACCCTATC GCGGCACGTT ATGACCTATC	26100
TTGGAAAGAA ATCTTCAAA AAGGATTGAT GTTGGCTAGT TTTAACTTTC CTGGTTCTT	26160
CCTCATGTTA GCCATTCTTG TCCTCATTGT GATGGTCTT TATCTGTCCG CCTTCAGTCT	26220
ACTCTTAGGT GGCTCAGTCT TCCTACTTTT TGGGTTTGA CTATTGGTCT TTATCCAGAC	26280
TGGATTGATG GAGAAAATTT TCGCAAAATA CCAATAGGAG CTTTATTTCT GAAACTACTT	26340
TCAAAGGCTC CAAACGCTAT TCTATAAGCG AGAACTAAA ATCGG	26385

173

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2716 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CCTGCCCGCA TTGCCCTAGG CATTAAGTAA ACATATAAAA GCATGTGAGA GACTGTTGGA	60
AAAGCGAGGA AATTTCCCTT CTTTTCCTCT AGTCTCTCTT TTCTTTTGCT GATTTTATTC	120
AAAGAAAATG ATATAATAGT AGTTATGGAG AAAAAGAAAT TACGCATCAA TATGTTGAGT	180
TCAAGTGAGA AAGTAGCAGG ACAGGGAGTT TCAGGTGCTT ACCGTGAATT AGTTCGTCTT	240
CTTCACCGTG CTGCCAAGGA CCAATTGATT GTTACAGAAA ATCTTCCAAT CGAGGCAGAT	300
GTGACTCACT TTCATACGAT TGATTTTCCC TATTATTTAT CAACCTTCCA AAAGAAACGC	360
TCAGGGAGAA AGATTGGCTA TGTGCATTTC TTGCCAGCTA CACTTGAGGG AAGTTTGAAA	420
ATTCCATTTT TCTTAAAGGG AATTGTGAAA CGCTATGTAT TTTCTTTTAA CAACCGGATG	480
GAGCACTTGG TTGTGGTCAA TCCTATGTTT ATTGAGGATT TGGTAGCAGC TGGTATTCCA	540
CGTGAAAAAG TGACCTATAT TCCTAACTTT GTCAACAAGG AAAAATGGCA TCCTCTACCA	600
CAAGAAGAGG TAGTCAGACT GCGCACAGAT CTTGGTCTTA GTGACAATCA GTTTATCGTA	660
GTAGGTGCTG GGCAAGTCA GAAACGTAAA GGGATTGATG ACTTTATCCG TCTGGCTGAG	720
GAATTGCCCTC AGATTACCTT TATCTGGGCT GGTGGCTTCT CTTTGGTGG TATGACAGAT	780
GGTTATGAAC ACTATAAGAA AATTATGGAA AATCCCCCTA AAAATTGAT TTTTCCAGGC	840
ATTGTATCGC CAGAGCGGAT GCGCGAATTG TATGCTCTAG CGGATCTTTT CTTGTTGCCT	900
AGTTACAATG AGCTCTTTCC TATGACTATT TTAGAAGCTG CGAGTTGTGA GGCTCCTATT	960
ATGTTGCGTG ATTTAGATCT CTATAAGGTG ATTTTGGAGG GAAATTATCG GCGACAGCG	1020
GGTAGAGAAG AGATGAAAGA GGCTATTTTG GAATATCAAG CAAATCCTGC TGTCTTAAAA	1080
GATCTCAAAG AAAAGGCTAA GAATATTTCC AGAGAGTATT CTGAAGAGCA TCTGTTACAA	1140
ATCTGGTTGG ACTTTTATGA GAAACAAGCC GCTTTAGGGA GAAAGTAAAA AGTGAGGTAA	1200
TCTATGCCAA TTGGTTTATT TACAGATACC TATTTTCCTC AGGTTTCTGG TGTTCGACC	1260
AGTATTCGAA CCTTGAAAAC AGAACTTGAA AAGCAGGGAC ATGCTGTTTT TATCTTTACG	1320
ACGACAGATA AGGATGTCAA TCGCTACGAA GATTGGCAAA TTATCCGCAT TCCAAGTGTT	1380

174

CCTTCTTTG CTTTAAAGGA TCGTCGCTTT GCCTACCGAG GTTTTAGCAA GGCACCTGAA	1440
ATTGCTAAAC AGTATCAGCT AGATATTATC CATACTCAGA CAGAAATTTT CTTTGGCCCTG	1500
TTGGGGATTT GGATTGCGCG TGAATTGAAA ATTCCAGTCA TCCATACCTA TCACACCCAG	1560
TATGAAGACT ATGTCCATTA TATTGCTAAG GGGATGTTGA TCCGGCCGAG TATGGTCAAG	1620
TATCTGGTTA GAGGTTTCCT GCATGATGTG GATGGGGTTA TTTGCCCTAG TGAGATTGTC	1680
CGTGACTTGC TATCTGATTA TAAGGTCAAG GTTGAAAAAC GGGTCATTCC TACTGGGATT	1740
GAATFAGCCA AGTTTGAGCG TCCGGAATC AAGCAGGAAA ATTTGAAAGA ACTGCGTAGT	1800
AAACTAGGGA TTCAAGATGG TGAAGAGACG TTGCTTAGTC TTTGAGAAT CTCCTATGAA	1860
AAAAATATTC AAGCAGTTTT AGCAGCCTTT GCTGATGTTT TGAAAGAGGA AGACAAGGTT	1920
AACTGGTAG TAGCTGGGA TGGCCCTTAT CTGAATGACC TCAAAGAGCA AGCCGAGAAC	1980
CTAGAGATTC AAGACTCAGT CATCTTTACA GGGATGATTG CTCCTAGTGA GACGGCTCTT	2040
TACTATAAAG CGGCGGATTT CTTCAATTCG GCATCGACAA GCGAAACGCA AGGTTTGACC	2100
TACTTGAAA GCTTAGCCAG TGAACACCT GTCATTGCTC ACGGAAATCC TTATTTGAAC	2160
AACCTCATCA GTGATAAAAT GTTTGGAACC TTGTACTATG GAGAACATGA TTTGGCTGGT	2220
GCTATTTTGG AAGCCCTGAT TGCAACACCA GACATGAACG AGCATACCTT ATCAGAGAAA	2280
TTGTATGAGA TTTCAGCTGA GAACTTTGGG AAACGAGTGC ATGAGTTTTA TCTGGATGCC	2340
ATTATTTCAA ATAACCTCCA GAAAGATTTG GCTAAAGATG ATACGGTCAG TCAGCGTATC	2400
TTTAAGACAG TTTTGTATCT TCAGCAACAG GTGGTTGCTG TACCTGTAAA AGGATCTAGA	2460
CGCATGTTGA AGGCTTCAA AACACAGTTG ATCAGTATGA GAGACTATTG GAAAGACCAT	2520
GAAGAATAGA AAGAGGAACA GCTATGAAAA AAACAATTAA TGAGAAGCGG TCGTGATAAA	2580
AAGATTGCGG GTGTTTGTGC TGGGGTGGCC CATTATCTGG ATATGGATCC GACTATCGTT	2640
CAAGTCATTT GGGGTGTTCT TACTTGCTGT TACGGAGCTG GAATTGTAGC TTACATTATT	2700
TTATGGATTA TCGCGA	2716

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13926 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTTTGGTTTT GCCTTATTCA AGACATGAGG GCCATCAGGA ATGATCTGAA ACTGCGAATC

60

175

TGTTAACAGT CTATGGAGAG CTTTCATAGA ACTAAGATTC GGTTTATCTT TGCTGCCACA	120
AATTAGTAAG GTTGGATAAG GGTAAGTTCC TGCTATATCC GTTAAATCAA GTGTCTTCAA	180
CTCCTCAGAA ACTCCGACCA TAAGAGTCTT GTCTGCTCCC TGTTTTTCAA ATACTCTTTT	240
GGGAAGTAGT TTAAAAATCA GCAATTGAAG ATAAATAGG ATATTCCCTG CTAATTTAAG	300
CGGGCATCCT GACAGAATCA AAGCTCGAAG ATTTGGTAAA TCGTAACTGG AAAGTTCTAG	360
TGTCAGGGCA GCACCTAAGG ACAATCCAAT CAAAACAAA GGTTCTGTCT CTTGAGCTAG	420
GTGCTGATAA ACTCGCTCTT TAGCTTGTG ATAGTTACTA ACTCCAGAAG GAAATAACTC	480
GATAGCCTCA GAAGGATAAT CTGTCACTAG ATTCCGAAC TCTTTCCAAG ACTCTGTGTA	540
CTGCCCTAAC CCATGCAAAA ATATTAATTT CATCTAGTTC TCCTCAAGGC TTAATTCATA	600
CAAGCCTCTC ACTGCATTAC AGCCGTAAAT AGCTTCTGCT TGGGTAAAT CTGCCAAGGT	660
CAAGACTTTC TCTTCTACCT GTCCTGTTTC TAGCAAATGC TGACGGTAAA TTCCTGGCAA	720
GATTCCAAGT CGGATAGGCG GTGTGTAGAG TTTCCAGCG ATTTTCAGAA CCAAATTTCC	780
TATAGAGGTT TCAAGCAGTT CTCCTGACTT ATGTGTGTAA ATCTTCTCTT GTTCTCTAG	840
GCTCAAATGC GGTCGGTGAG TGGTTTTAAA GTAGGTAAAG GATTGATTCA AAGCAGCTTC	900
CTGAAGACAG ACTTGGGCCT GACAAAAGCT TGTACTGAGA GGGGTAAATA CTTGACGATT	960
GACTTCTATC TCTCCAGATT TGCTAAGGCT GATTGCGAAG CGGTAATCTC GATTAGCTTC	1020
ACAATCCTGA CACTCTTCCT CAATCTTGTG TCCCAAGTCT TCTGCATCAA AAGGAAAAGC	1080
AAAATAACGA CTAGCTTTTC TCAGCCTTTC CAGATGTTGT TCTTCAAACA TCAGTTGTTT	1140
TTGGCTGATT TTTCCAGTTG TAATTAATTG GAAGCGAGCT TGTTTACGAT AGAGAAGTGC	1200
TGCCTTTTGA TGAACCTCTC GGTATTCAGA TTCCCATGTG CTATCCCAAG TAATCCCTCC	1260
GCCAACTCCA TAAATGGCTT GACCTTTGTG AAGTTGAATG GTACGAATGG CCACATTAAA	1320
AATCCGTCGT CCATTTGGAA GCAAGAGACC AATCGTTCCA CAGTAGACTC CACGCGGTTG	1380
AGGCTCCAAG TCCTTGATAA TCTCCATTGT CGCAATTTTC GGTGCACCCG TTATGGAACC	1440
ACAAGGAAAG AGTGAGCGGA AGATTTCAAC AAGGTCCACA TCCTCTCGCA ACTGACTCTT	1500
GATGGTCGAA GTCATCTGCC AAACAGTTGA ATACTGCTCT ACCTGACACA GACGCTCCAC	1560
GTGCTCGCTC CCAACTTCAG AAATACGGTT CATATCATTG CGCAAGAGGT CCACAATCAT	1620
CATATTTTCA GAGCGATTTT TGGGATCCTG TTCCAACCAA CTGGCCTGTT CAAGATCTTC	1680
TTGGTCAGTT ACCCCACGCT GAGTCGTCCC CTTCAATTGGT CGTGTGTGCA ACTCGCGATC	1740
ATTTTGCTCA AAAAAGAGCT CTGGGCTCAT GGAAATCACT GTCATCTCGT CATGTTCCAC	1800

176

ATAGGCATG TAGCCCGCCT CCTGCTCTAC CACCATACGA TTGTAGATGG CAAAAGGATT	1860
GGCATTAAAC TTTTGCTTAA GTTGGACGGT GTAGTTGACC TGATAGGTAT CTCCTTGCCG	1920
TAAATGATGG TGAATTTGGG CAATGGCCTT TTCATAGTCT GCTGCAGACG TTACTTCCTG	1980
CCAATTTGAG GGCAAATCAA TATCCTCATA AGTCAGAGGA ATAGGGGAAG TTTCTACGAT	2040
ATCATGAACA GTAAAGTAAA GCAGGTACTC TCCAGTAGG GGATCCTTGT GAACTGCTAA	2100
TTTTTCCTCA AAAGCAGGTG CAGCCTCGTA GCTGACATAC CCCACCACAT AATAACCTTG	2160
CTCTTGGTAG CTTTCCACTT GTGCCAGCAA ATCTGCCACT TCTTCTACAT TTCTCGTTTT	2220
CAACTCTTTA ATAGGCTGGG TAAAGGTATA TCTCTCCCCC AAAGTCCTAA AATCAATCAC	2280
TGTTTTTCTA TGCATACCTT AAGTATAGCA TAAAATAAGA AAACCCTCAT CCGCAAAGCA	2340
GATGAGAGAT TTCAATTATT TAAAGATTGA AGTTTTAAAG CTATTTGTTT GTTGAAGAAG	2400
TTTCTTATAA ACAGCTTCTT TTAATTTAAC TGTATTATTC ATAGATACTG TTTTATTACC	2460
GTTTGCTTCT TGTTTAAGAG TTTCCGCATC TTTTTTAAAC GCTTCTTTAA ACAATGTCAG	2520
TAAATCATCG TATGATGAAA CGGAAGAACC ATTTACTTCG AATGTTGTTA ATCCTTTCGT	2580
TGCTTTATCT TTAACCTCTT TGAAGTAAGC TTTTTTAAAT TCTTCAATAG TATTAAATGT	2640
ATTGTTAGAT ATTTTCTTGA TAATATATTC ATCACTTAGA ACAGACTCAC CATCTGTTTT	2700
AGATTGTTGT TTATATTTAT TTGAAGCATA ACCTAAGAAC CCATTTTCGT ATCCGTAGTA	2760
ACCCCATAAAT CTAAAAGCAT TATGTTTGAA TGAACAGCT CCAGGAGCAC CTTTACTAGT	2820
ATTACCTCCG TAGATACCGG TCATCATTCT AACACCTACA TAAGGTGATT GATCGTTATA	2880
GCTAATGCT TCGGGTTTAT AGATACCATT ACCTGGATTG CGATTAGTCA TTAATGTTG	2940
ATCAACTAAA TCATTAACAG ATTGAATATT TAATTCATTT TTCTCTTCTT GACTTAGATT	3000
TCGAATTTTA TCCCATTGAT TTAATTTATT GTTATCACGG TATTCTCTAT CTATTTTTTT	3060
GAACCATGCA CTATTTAAAT CTTTATTTTG TTGAGAAATC ACAGATTCAG CCTCAATTTT	3120
ATCAAGAAGA GTTAAAGTGT CATTATAACC CTTTCATATAT CTATTAATAT CTTCTCGTGT	3180
TTTTAGAGTT TTTGGATCTG TAATATACCA CTGATTCCTA TCATTTTTCG GTTTAAATAC	3240
CATATTAATA CCTAAAGAAC CAACTCATC AAATCCACTA CCAGTAACAG GAGTTTGTAG	3300
CATACCCTGA GCATATGCTT CAGCATCAGT ACCTTCACGG TGTCCAAAGC CACCTAAGTA	3360
AATCGCACGG TCGTTGACGT GTGTTGTTTC ATGTGTGTAA ACTGAAATAC CGTATTCACC	3420
AACCATTTCT AAATGAACAT ATTTTACATC AGTTCTAATA TCATCAGAGT TAGGATATAT	3480
AGCAGCATAA GTCCTGTTC CATTATAATT ATAATACTTA TCCATAGGAC CAAAGAATTC	3540
TCTAAGAGGA GTATATACTT TGTCGGTATT ATAGCGGCCA TATTTTTCAA CCCATCCACC	3600

177

AGGAGCGTTA TAACCTTCCC AAATAGGAAT AACAGCATCT CTTAGTAGTC GTTGTTTAAC	3660
GTTATCAGAC GCTAGACGAT ACCAGAAATC ATAATAGTTT CTATAACCAT CTGCAGCTTT	3720
GTTAACGATA TCTTTAATAT CTTCTAATGA TTTTCTACCT AATCGCTCTG CACTACCAAA	3780
GGCAATTGCA TTATAATTTG AAATPAAATA AAGATGTGCT TTATCAATAT TCAGTAGTGG	3840
GAGTATAGTA TTTCTAAGGT GACTTCGTTT TAAATTATCG AATGCACGAT GTTTAGAATT	3900
TTTAATTTCT TCGACCTCAG AAGCGCGTTC TGCGATGTAG ACATGGTCTT CTGTAGCATC	3960
AATAAACCAA TCGTTCATAT TGTCTATATT TGTGAACAAT TGTCTATTAT AATTTAAAAA	4020
TGCATCTAAA TTACCTGATT TAGTATATTT AGCCAATACT TGACCGAATG CGTCGAATGT	4080
ACGTGAACCT TTAATGTTGT TCTCTTTAGA ACCGATTTCA ATTAATCTGT CTAATACGCT	4140
AACTTTTTCA CCATAGAAAT CTGGTTTGAA TAGCATTAAT TCTTTAATAT TAACATCACC	4200
AAATTTAACT CCATAGTAAC GATTTAGGTA AGTTAAACCT AGTAATAAAG CTGCTTTGTT	4260
TTTCTCGACT TTATCACGAA TCATTTGACG AGCAGCTGGA GAATCATTTA GTTGATGTTT	4320
TTCTGTTTGA ACTAATTTTG TGATTAGGTT TGTTAAGTTT TCTTTAACAT CTGTGAAGCT	4380
TTCTTCTAAA TATAAATCTT TGATTGCATT AACTCTATAG TCACCTAATC GATTTAGATG	4440
CTGATACATC GTTTGAGACT GAAGCTCTAC TGATTCTAAA ATAGATTTTA TATCATTAAC	4500
AAGAGTAGTG TTATCTTTT GAACGATATT AGGTGTATAT TTAATTCCTA AGTCAGTTAT	4560
AGTATATTCT TTTACATTAC TTAAACCTTC ACTGCTAGAA GACAAGTTAA AGTAATCTTT	4620
TGTACCGTCC GCATAGTGAA CAATAATTTT ATTAGCTTCA TCTAGGTTTG TGATAAACTC	4680
ATTGTTGTTT ATCGCGGTAA CAGAAAGAAC TTCTTTAGTA TTTAGATGGT GTTCTTTATT	4740
TAATTTATTA CCTTGATATA CAATATAATC TTTATTGTAG AATGGTATTA ATTTTCAAG	4800
ATTTTATAG GCTTGGTTAT ATTCAGCGTT ATAATCTTGA ATACTAGAAT AGGCTTTTTC	4860
TTCAATTAACT TTTGCAAGAG GAGATAGATC ACTTTCTAAT TTATCAGCAG TAATATTGAA	4920
AGTAGTAACT TTAGCATCAG CTTGTTCTTT AGTTAATTTA GTAAATGTTT TAGATTTTCT	4980
AAATGATCTA TTACCTGACG AATATCCCTC TACCGCATAT AAATCTTTTA TATGAGCACT	5040
AGCATAATCA GAATCATCAA CGTCGTTAGA GCCGAATAAC TCCTCTCCAC GGATAATCTT	5100
AGCATAGCTG ACAGAATTAC TTACCGTACC TACAGGCCAA GTCTTACTTG CTATTGCTCC	5160
AACTTCTACT GGATTTGAAA CATCTATTTT ACCTTTTACA ACCGACTCAG TTAGGAGAGC	5220
TTTTGTACCA ATAAGATGGT CTAGAGTTAA TCCATAATCT ACTTTAGGAA CTAACAAGCT	5280
GGCGCGTGT TTGTTTCCTG TAATAGTAGC ATCAACATAT GCTTTTCTAA CAATTCCTCT	5340

178

ATAGTTTGTA CCTGCAATTC CCCCTGTATG AGAGCCATTT CCACTTGTAG AGTGTAGTTT	5400
GCCAAAGAAA GCAACATTTT CAATACGAGT TCCATCATTC ATATTATTTA CAAATCCAGC	5460
AACATTATTA CGACCTGAAA GTGTGCCTGT AATTTTGACA TTTGTAATAA CTGAAGAACC	5520
TTTCATAGTA TTGGCTAATG ATGCAATATT ATCTTGACCA GAACGTTCTA TCTCTACATT	5580
TTCAAAATTC ACATTATTTA TCGTTGCGTT TGTATACACA TTAAATAATG GATGTTCCAA	5640
TTTCAGTAATA GCAAATTGTT TTCCTTCAGA ACTTAAAAGT TTCCTGTGA ATTCTTTAGT	5700
GATATATGAT TTTCCATTAG GAACAACATT TCTAGCGCTC ATTGATTGTC CCAGACGATA	5760
TTCTTTTGAA GGATCGTTTT GAATAGCTTC CACTAATTCT TTGAAATPAT AATATACATT	5820
ATCTTCGTGG ACTTTAGGTT TTTCAATATA GTGAACGTAT TCTTCTTCAA ATTTATTATC	5880
AGCAGTTCTA GAGACTAAAT TGTCTGCGAT TGCTGTAAC TTATATACAG GTGTTCCGTT	5940
AACCGTAGTT TCTTCTATAT TTTTAACAGC TAGTAATGTA GTTTTCTGAT TATTTGAAGT	6000
TATTTTAAA TAATAATTGC TCTTATCATC AGGAATAGTT GTTATCAGTG ATTCATTAGT	6060
TTCTTTTCCA TTTTCGTATT TGATTAAATC TGTACGTTA ATATTTTAA GCTCAACTTT	6120
TTTAAGATCT AATTGAATAT TTTGATTTTC TAGAGTTTCA GTTCTTTCAC CGTTACCTCT	6180
GTCGTAAATC ATAGTTGTAG ATAGGGTGTA TTCTTTGTAG TACTCTAGGT TCTTAAATGC	6240
AGCGCTTATA GTTCTGTGTT TTACCTGTGC ATCTGTAAGG ACTACAGTAT TAATAACTTC	6300
TTCTCCTTTT TTCAATTCAG CTGTGATTGA TTTGATTTT GTTTTGTGTTT GATTTTCTAG	6360
AGTATACTTA GCAACAGCTT CACGTTCCAA TATTTTCTTA TCGGTACTAG TCAATGTTAA	6420
TATTGGCTTT TCAGATAATT CAACCAATTT TTCAATAGTT GCAGTTAATT TTTCAACAGC	6480
TTCGTTAACT TCACTTTGTT TAGCATCTGT ATTAGCTGCA ACTTTTTCAG CCTTTGTAAC	6540
TTTCAGTTTG AGGTTTGGCC AACTTCTATC ACTGTAATGT TCTTTTACCT TTGTTTTTGC	6600
ATCTGCAATC GTATTGTTTA ATTCACTTTT ATCAACGTTT AGACCGTCAA TAGCGGTTTT	6660
AAGTTTATTT GTCTCGCTAT TTACCTCAGG CTGTTTACA GGCTCTGAAG CATAGACACC	6720
TTTTGCAGTT TCTAAAACAG GTCCAAGAGC ATTGTAACTT GCTGTAGAAT AATCAGTAGG	6780
AGAAACTGAA CTAGCTTTAT CAATTTGATT ATTTAACTCA CTTTATCAA CTGGTCTTTT	6840
AGTACCAATA CCCTTTATTT TATCTTCTGG TTTCGGTGTT TCCTCTACAG CCTTCTCTTC	6900
TTCAGGAACT TCTGGTTGCT TTTCTGGCTC AACTGGTGCC GTGGTGCCT GTTCGTCCTC	6960
TCTTGGCGCG ACTGGTTCAC CTGCTTGTTT AACTTTTGGT TCCTCTGTTG GTTCTGTTTG	7020
TTTTTCTACA GCAGGCGTTT CAACTTTTGG TTGTTCAATA GATTGATTAA CAGTCTCCTC	7080
TTTTGGTCTT ACAGTTTCTT CAGCCTTGGT ATCTGGAGTT GACTCTTCTT GTTTCGGTGT	7140

TTCTCTACA GCCTTCTCTT CTTGAGGAGC TTCTGGTTGC TTTTCTGGCT CGACTGGTGC	7200
CTTTTCGTCT TCTCTTGGCG CGACTGGTTC ACCTGCTTGT TCAACTTTTG ATTCCTCAGC	7260
TGGTTTGTCT GATGGTTGAC TTCTGGCTT AACTGCTACT TTTTCTCTG GTTTTGACTC	7320
AACTTCTCCA CCTACTTCTT CAACTGGAGC TGCTTCTGCT GAATCTTCTT TCCCCTCTTC	7380
TACTTTAGGA AGGGTGTCTG CAGTAGGTTT TACCTCCGAT TTTGGTTCTT CCTTTGGACT	7440
TTCTTCTGTT TTAGGTGCTT CTCTTTTGG AGCTTCCTCT GTCTCTACTA CTGGTTTTC	7500
TGCTCTAGCT TGCTCCTGAT TTGTTATTGA TTGAGGAGTC TCAACTTCGA CCACAGTCAC	7560
CTCTCCAGGT TTTGCTGAGG TTTCTTCTAA AACAGTGTCC AAGCCAAAGC TTTTGAGGAT	7620
GTCACCTGAT AGATAACCA CATAGCGATA GCCCTCCATT TCAACAACAC CCTCTCGACT	7680
AGCCAGCGCT AGGGTCGCAA CTGGGTCTAC AGCCCTGCA CTAGGAAGAA CTACCAATCC	7740
CATAGCTCCA ACTAGAAAGA CGCTAGCAAT TTTCTTCTC TTGTAGATTA AAAGCAAGCT	7800
CCCAACAGTC AGCAAACCA AAGCTGTCAA AACAGATGCT TCTGTCCCTG TTTGAGGCAA	7860
CTGATCTTTT TGATACACCA AACCATATAC AACTTCATT CTGTGAGGCT TTCCTGTCTG	7920
AATTAAATCT TTAGCTTCTT GTGAAATAAT CTCTTTATTT ACATAGTGAT AGGTGGCTGC	7980
GTCCACTACA GAAGGAGCCA TCAAAGGCT TCCAAGAAAT ACAGAGCCTA CAACTCCCTT	8040
AATCTTACGA ATTGAAAAAC GGTCTTTTTT AAACACTTTT ATCTCCTTTA TTCATTCTCA	8100
AACTTCCTA ATAGCATCTT GCGATAGTG CGCACGCGCA CCTCCGATTA ATTTTGAGCG	8160
ACTAGCCAGT GCCGTACAT GGGCATGACC AATCTCTCTC AAAATAGGGC GAATCGGAAC	8220
CTGAACATGC TTGACATGCA TGCCAATTGC AGTGTCTCCG ATATCCAATC CAGCATGAGC	8280
CTTGATAAAT TCAACCTCAA CTGGATCCTG CATAAACTTA AAGGCTGCCA ACTGCCCCGA	8340
ACCTCCTGCA TGAAGAGTAG GATGGACACT GACAATTTC AGACCAAAT GCTCTGCCAC	8400
CTGACGTTCA ACAACGAGAG CCCGATTGAC ATGCTCACAA CCTTGAAGTG CTAAATGGAT	8460
ACCTCTACTA CCTAGAATAT CCAAGATAGT CTCCACTATC AGCTCACCAA TCTCTTGACT	8520
GGATTCTTTC CCAATATGAC CACCTAGCAC CTCAC TAGAA GATAGACCTA AAACAAAAG	8580
GGCCCCCTGC TTCAAATTGG TCTTTTCTAA AACATCTTCC ACTACCTGAC GTGTTTCTCT	8640
TTGAATCTGT GTCTCGTTCA TCTCTGTTAC CTCTGTTGTC ACTCTTCTAT CATACCGTTT	8700
TTTCTTGTTT TTAGCAAGAT AGACAACCTA GAAAGTTTGC CCAATTACGC ATAAACTCC	8760
CAGAATTGAC TGGGAGTTAG CTAGTTTCTA TTCTATTAT ATATATTTC ACTTTCGTCC	8820
CTTTTGGGG TCTAGAATCA ATCTTCATAT GGTAAATTGGC TCCAAAATGA AGTTTGAGCC	8880

180

GTGATCGAC ATTTTGAAGA CCAACTCCCC CACGTTTGAG TTGACTTTGA CTACTATCAC	8940
CAGCATCTTG GAAGCCAACG CCATCATCCT CAATACGGAT GACCAATCCC GAATCCTGTT	9000
TCTGGACAGA AAGTTTAATA TGGCCCTGAC CTTCCTTTTC CTTAATGCCA TGGTAAAGAG	9060
CATTTTCTAC AAGGGGTGT AGGACCAGCT TGGGTAAGAC TAAATTATCA AAGGCAACAT	9120
TTTCATTAAT TTCGTATTCC AGCTTATCTC CATAGCGTTG TTTCTGGATA AAGAGATACT	9180
GGCGGACATG ATTGATTTTCG TCAGAGAGAC AAATCAAGTC CTTCCTTGA TTGAGCGCCA	9240
AGCGGAAATA GGTGCCAAG GACTTGGTCA CCTGCACCAC TCGCTGACTA TCATGAAATT	9300
CAGCCATCCA GATGATGGTG TCCAAAGTGT TATAGAGGAA ATGTGGATTA ATCTGGCTCG	9360
AAAGGGCTTG AAGTTGGTAC TGACGGGTCG TTTCTTCCTG GCTACGAATA GCTACCATCA	9420
ACTGATCAAT CTGATCCAAC ATAGCATTA AATTGGCGAGT TACTTCTCTC AGTTCATAGG	9480
CACCAACTTC CTGGCACGA AGATTTTGAG CACCAGAAGC AATTTCCAAC ATGGTTTCTC	9540
TCAAATCCTT CAAAGGAGCA ATCCAGCGTT TAAGACTGAA CCACACTAAG CAGAGACAGA	9600
CAAGAAGAGA TGTGACACTG GCCCCAAGCA AGGTCCACAA GAGCTGACTC CGAACCTGGT	9660
CTAACTTTTC CAATGATGAC ACGCCAAGCA CCGTCCAATC AGTTCCTGCA ATCTTCTCTT	9720
GACTGACGTA GGATTTGTGA CCAGGAGTAT AACCTGACC TGTATCGATG TAGGTTTCA	9780
TAGCCTCCAT TTGCTAGAC GAACTATAAA CTGTGTGTTG AGGATGGTAG ACAAATTCAT	9840
GGTTTTCATT GATAATGAAG GCAAAGCCCT GCTGCCCAA CTGGAGTTGA TTGAGATAGG	9900
CTTCCAGAGT TTCATAAGAA ATATCCAAAC GAAGCACACC AAGATTGGCT CCCTTTGCAT	9960
CAACAAGTTC TTGAGTGACA GAAATGACCC ACTGACTATC TGATTTACGA GCTGGAGTCA	10020
AAACAGGCAT AGCTCCCTGA TGAATGGCCT TTTGGTACCA ATCCTCAGCC ATCATATCAG	10080
AGGAAGTTT CATCTGCACA CTGTCATCTG TAGAAATGAC CTGACCAGAT TTGGTCACCA	10140
GCACAACAGT TTTCAAGTCC TTATCTGACT TCAAGATGGT CAAAAACAAA TCTCGGATTC	10200
CCTGCACCTT GTCTTGACTG GGATTCTCAG CATAGGCCAG AACATCCGTC TGCTGGGTCA	10260
AACCAGTCGA GGTGGTTTCT AGTTTTTTGA TATAAGACTG AATAAAGTGG CTAGTCTGGC	10320
TGATGGTCGT TTGGCTGTTG CCCTCAATGG TGGCCTCAAT GGCTGAAGAA CTTGATTTGAT	10380
AGTAGAAAGT TCCAACCAGA GCTAGGAGAA TGAGAAAGAC CAGAAAGATG GAAATAACCA	10440
TTCTAACTAA AAGAGAAGAA CGCTTCATCG GTCTTCTCCC TTCTTAAACT GACGAGGTGT	10500
CACACCTGCA ATCTGCTTAA AACGTTGGGT AAAATAGTTC ATATCTTCAA AACCAACCTT	10560
CTCTGCGATC TCATAAATCT TCAGATCTGT AGTTAAAAGC AAGAGCTTGG CTTGTTTAAAC	10620
ACGTTCTCTC ACCAGATAAT CCTGAAAAGG CAAGCCCAAC TCTTTCTTAA TCAAGGAACT	10680

181

CAGATAGGTC GGA	CTAAAAC	CTAAGTCACT	GGCTAAAGAC	TTTAAACTAA	ATTGGCTATC	10740
AGCCAGATGA GACT	GGATTT	TCTGGGCCAT	GTTTCCTTCA	AACCTATTAG	TCAATAAATC	10800
TTGTAAGTGC TCT	TCTTCT	CTTCCTTGTC	TAGTTTTTGT	TTGATTTTCC	CCAACATTTC	10860
CTCAATATCC TGAC	GAGAAA	AGGGTTTGAG	CAGGTAGTCG	TCCACACCTA	GTTTGACAGC	10920
AGACAAGGCA TAAT	CAAAAT	CATCGTAACC	TGTTAAAAAG	ACCAAATGAA	CCTGAGGATA	10980
GGTTTCTCGT ACC	CAGACTGG	CCAACTGGAT	GCCATTTAGA	TGAGGCATGT	TGATATCGGT	11040
TAAAATGATA TCT	GGCACCT	GCTTTTGAT	CAATFCCCAA	GCCTGCCTTC	CATTTTCAGC	11100
CTGACCGATG ATTT	CCATAT	CGTAGGCTGC	TACATTGACC	AGTTTAGTCA	AACCTTGTCT	11160
TACCAGATAT TCAT	CTTCTA	CGATTAAGAT	TGTGTAGGTC	ATGCTCTGCT	CCTTTACCAC	11220
TTACTAGTAT CAG	TATAGCA	AAATTCTCCT	CTAACTGCTT	AGGAAAGACC	TCTTATACTC	11280
AATAAAATC AAAA	AGTAAA	CTAGGAAGAT	AGCCACAGGT	TTCTCAAAGT	ACCGCTTTGA	11340
GGTTGTAAAT AAA	ACTGACG	AAGTCGACTC	AAAGTATAGC	TTTGAGGTTG	TAGATAAAAC	11400
TGACGAAGTC GAT	AACCCTA	CATACGGTAA	GGCGACGCTG	ACGTGGTTTG	AAGAGATTTT	11460
CGAAGAGTAT TAAT	CAACAT	AATCTAGTAA	ATAAGCGTAC	CTTTTCTTTC	CATTTGGTCT	11520
TTGGAATAA AGCG	GATAGA	GAGGCTATTG	ATACAGTAAC	GTAAGCCGCC	CTTGTCCTGT	11580
GGACCATCCG TAA	AGACATG	CCCAAGGTGA	GAATCTCCTA	CTCGGCTCCG	CACTTCCATA	11640
CGCGTCATAT TGT	AGACTT	ATCTTCCTTG	TAGGTGACAA	CATCTGGACT	GATGGGTTGG	11700
GTAAACTAG GCC	AGCCACA	ACCAGACTCA	AATTTGTCTT	TTGATGAAAA	GAGAGGTTCC	11760
CCAGTTGCTA TAT	CCACATA	GATACCGGAT	TCAAATTTAT	CCCAGTAACG	GTTTGAGAAA	11820
GCTCGTTCTG TTT	GATTTTC	CTGGGTAAC	GCATACTCCT	CAGGTGACAG	GGTCTTTTTC	11880
AATTCCTCAT CACT	TGGT	TTGATATTG	CTGGCATCAA	TGACAGGATA	GGCCGCCTGA	11940
TTAACATTGA TAT	GGCAGTA	GCCATTTGGA	TTTTTCTTGA	GATAGTCTTG	ATGGTAATCC	12000
TCAGCCACCA CAAA	ATTCTT	CAAGTTTTC	TTTCAACTG	CTAGAGGTTG	ATCGTATTTT	12060
TTAGCCACCT CAT	CAAAGAC	TTGGTTAATC	ACTTCCAAAT	CCTTGTCTATC	TGTGTAATAA	12120
ACACCAGTAC GGT	ACTGGGT	CCCCACATCA	TTTCCTTGTT	TATTTTGTCT	GGTTGGATTG	12180
ATAATGCGGA AAT	AGTGAAG	CAGGATTTCC	TTGAGAGAAA	TTTGCTTGGC	ATCATAGGTG	12240
ACATGGACGG TTT	CTGCATG	ACCTGTTTGG	TTAATCAATT	CGTACTTGGT	TGTTTCTCCT	12300
CTACCATTTG CAT	AGCCTGA	AACGGCATCC	GTCAACCCCG	GAACACGTGA	GAAATATTCC	12360
TCCACTCCCC AGA	ACAACC	TCCAGCTAGA	TAAATTTCTG	GCAAGTCTGC	GTCTTTACTA	12420

182

ATTCTGTTT TTTTCACTGC TTTTCCTCCT TGGCTAACTG CCGCCTTTTC AATTTGCGAG	12480
GCATCTGTCT GCCCTGCATT TCGTATCAAT AGAACATAGA AACCGGTTAT GGCTAGAAAA	12540
AATACTCCTA GCAACAAGAA GATTTTAAAC TTATCATTCA TAAGACGCCT CCTAGGCTAA	12600
TTCCCTCAAA GTTTGCAAAA TTGCATCTTT TTCCATGAAT CCTGGATGTG TTTTGACCAG	12660
CTTGCCTTCT TTGTCTATAA AGGCTTGGGT TGGGTAAGAA CGGACACCAT AAGTTTCCAA	12720
AAGTTTGCTT GATGGGTCAA CTAGGACTGG GAGATTTTTA TAATCCAATC CCTTATACCA	12780
ATTCTTAAAG TCCGCTTCAG ATTGCTCTCC CTTATGTCCT GGTGACACTA CTGTCAAGAC	12840
CACATAGTCA TCACCAGCTT CTTTAGCAAT CTCATCCGTA TCTGGAAGAC TAGCCAGACA	12900
GATGGAACAC CAAGAAGCCC AGAATTTGAG ATAGACTTTC TTGCCCTTGT AATCAGATAA	12960
ACGGTAGGTC TTGCCATCTA CTCCCATCAA TTCAAAATCA GCCACCTCTT TCCCTTTAGC	13020
TGCGCTTGTT TTAGTAGCTG TCTGCTCCGT CTTCAATTCA TCTTTCGTTT GGTGTTCACT	13080
AGTCACGGAC TTGCCTGAAC AAGCCGTCAA ACAAGGAGC GAACCTGCTC CAAGAACACA	13140
TGTTTGCCAT TTTTTCATAT TGATATTCCT TTCCATTTTA TTCAAATAAT TGACTTAAAA	13200
TTGAAGCATT TCCAAACAGA ACCAAGAAGC CCATCACAAT AATGAGAAAA CCACCCACTT	13260
TTTTGAGGAT TCCGAGATAG GGATGAAGTT TTCGAAATG TTTCAAAACA TAACTAGAGG	13320
TCAGAGCTAG AAGCAAGAAT GGTAGCGCCA AGCCCAGCGT ATACACCAAC ATGAGACCAG	13380
CTCCCTGCCA AGCTCCTGAA CCACCTGAAG CCGCCAAGGC CAAAACAGAC CCCAGAACCG	13440
GCCCCACGCA AGGCGTCCAA GCAAACTAA AGGTCAAGCC CAATAAAAAAT GCCTGACTAT	13500
AGCCCTTACC ATTTTGCCCC TGTCTTGCA GTTGTAGCCT CTTTTCCTTA TAAAGCCCCT	13560
TAAAGTGTAG AATCTCCATT TGGTGCAAAC CAAGAAGGAT AATAATTGCC CCAGTAAGAT	13620
ATTGGAACCA AGAAGCATAA AGCAAATCGC CTAAAAAACC AGCTCCATAG CCCAACAAAA	13680
TAAATATAAA GGAAATTCCT GCTATAAAGG CCAGAGTTCG TAATAAACTA GTAAGTGA	13740
TTGAAAATTT GCCGCTAGAA GCCTGAGCAC CATCCTTATC ATCTAGTAAC ACTCCTGTAT	13800
AGACCGGTAA CAAAGGTAAG ATACAAGGAG AAAAGAAGGA TAGAATCCCT GCCAAAAAGA	13860
CACTTAGAAA AAAGAAAATA TGACCCATAA AGTTCCTCCT ATCATTTTAT TGATAGATT	13920
ATTATA	13926

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20199 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCCAGCAGAA AAATGGCATT TGGAGATAAT GGAAATCGTA AAAAAACTAT GTTTGAGAAA	60
ATAACCTTGT TTATCGTGAT TATCATGCTA GTAGCAAGTT TATTGGGAAT TTTTGCAACT	120
GCAATTGGTG CCCTCAGTAA TCTATAAAAT AGATTCAAGA AAATTTAGTG ACTGGGATTT	180
CCCAGCCCTT TTTTAAAGTG AGAAGAAATA ATGAGTATGT TTTTAGATAC AGCTAAGATT	240
AAGGTCAAGG CTGGTAATGG TGGCGATGGT ATGGTTGCCT TTCGTCGTGA AAAATATGTC	300
CCTAATGGAG GCCCTTGGGG TGGTGATGGT GGTCTGGAG GCAATGTGGT CTTCGTTGTA	360
GACGAAGGAC TACGTACCTT GATGGATTTT CGCTACAATC GTCATTTCAA GGCTGATTCT	420
GGTGAAAAAG GGATGACCAA AGGGATGCAT GGTCTGGTG CTGAGGACCT TAGAGTTCCA	480
GTACCACAAG GTACGACTGT TCGTGATGCG GAGACTGGCA AGGTTTTAAC AGATTGATT	540
GAACATGGGC AAGAATTTAT CGTTGCCAC GGTGGTCGTG GTGGACGTGG AAATATTCCT	600
TTCCGACAC CAAAAATCC TGCACCGAA ATCTCTGAAA ATGGAGAACC AGGTCAGGAA	660
CGTGAGTTAC AATTGGAAT AAAAATCTTG GCAGATGTCG GTTTAGTAGG ATTCCCATCT	720
GTAGGGAAGT CAACACTTTT AAGTGTATT ACCTCAGCTA AGCCTAAAT TGGTGCCTAC	780
CACTTTACCA CTATTGTACC AAATTTAGGT ATGGTTCGCA CCAATCAGG TGAATCCTTT	840
GCAGTAGCCG ACTTGCCAGG TTTGATTGAA GGGGCTAGTC AAGGTGTTGG TTTGGGAACT	900
CAGTTCCTCC GTCACATCGA GCGTACACGT GTTATCCTTC ACATCATTGA TATGTCAGCT	960
AGCGAGGGCC GTGATCCATA TGAGGACTAC CTAGCTATCA ATAAAGAGCT GGAGTCTTAC	1020
AATCTTCGCC TCATGGAGCG TCCACAGATT ATTGTAGCTA ATAAGATGGA CATGCCGTGAG	1080
AGTCAGGAAA ATCTTGAAGA CTTTAAGAAA AAATGGCTG AAAATTATGA TGAATTTGAA	1140
GAGTTACCAG CTATCTTCCC AATTTCTGGA TTGACCAAGC AAGGTCTGGC AACACTTTTA	1200
GATGCTACAG CTGAATTGTT AGACAAGACA CCAGAATTTT TGCTCTACGA CGAGTCCGAT	1260
ATGGAAGAAG AAGCTTACTA TGGATTTGAC GAAGAAGAAA AAGCCTTTGA AATTAGTCGT	1320
GATGACGATG CGACATGGGT ACTTTCTGGT GAAAACTCA TGAACTCTT TAATATGACC	1380
AACTTTGATC GTGATGAATC TGTATGAAA TTTGCCCGTC AGCTTCGTGG TATGGGGGTT	1440
GATGAAGCCC TTCGTGCGCG TGGAGCTAAA GATGGGGATT TGGTCCGCAT TGGTAAATTT	1500
GAGTTTGAAT TTGTAGACTA GGAGACTGGT ATGGGAGATA AACCAGATATC TTTCCGAGAT	1560
GCGGATGGTA ATTTTGTTC CGCCGACAC GTTTGGAATG AAAAGAAAT GGAAGAACTA	1620

184

TTTAATCGTC TCAATCCAAA TCGTGCCTTG AGATTGGCAC GAACTAAAAA GGAAATCCA	1680
TCTCAGTAAA GAAGCTAAAA AATCCCGTGC CTCATCAGAC ACGGGATTTT GTGGTACGAC	1740
AGGCATGTAT AGCAAACCTGA ATCTGGAATA GCACAGCATA TCTTCTAAAA TATAGTAAAA	1800
TGAAATGAGA ACAGGACAAA TCGATCAGGA CAGTAAAATC GATTTCCTAAC AATGTTTAT	1860
AAGCAGAGAT GTACTATTCT AGTTTCAATC AACTATATTG TTATAAATTG ATTTGAATTT	1920
CAAAATTAAA TTGTTTGATT CTTATTTCAA TTTGTTATAG TATATCTGAT GTCAAAGTTC	1980
TCGGCGAGTC AAATAGCGAT TCCCAAGCCT GACTATCGTG AGGTAGCGGA TTAATATGGT	2040
CTGGGGATAG ACCGTTTAA GTCTGACGCT GGAAATAAGA ATTGTCAGAA GAAGGGATAG	2100
CGAAATCGTG GCTCTACGAA CAGGAACGTG ATAATAAGGC GTATATAGCG GATAAGAGGG	2160
CATCAAATC TAAAGTCCAA AAAGGTAGTC GTAACCTATA TCGGTAAATC ACGAGAGTAA	2220
TTGAATTCGT ACTAAGATTT TCTATTTTCA CTGTAACCTT TTAACGCCCT TATATCTTGT	2280
ATACACGAGG AAAGATGTAC GACTTATCCC GTGAGGTCTA TCACTATAAA GAGAAAACGA	2340
CAGATAGAAG TGATCCTGAG TCACGGTTAT CTGTCTGATA GGACGGTATG TATAAACGC	2400
TTCTGTGAAC TGAGAGAAGG GGGAGAAGTT CTTGCTAAAA TTAGTTGAA CAGCCGTATT	2460
CCGATACTTA GATAAGAGAT CTAGTCTTAG CTCCTACTCA GTTTTAGGGG ATAAAAAAGG	2520
GGCAATAGCG ATTCGAGAAA GATTATACTC TTCGAAAATC TCTTCAAATC ACGTCAATAT	2580
CGCCTTGTCG TATGTGTAGG ATACTGACTA CGTCAGTTCC ATCTACAACC TCAAACAGT	2640
GTTTTGAGCA ACCTGCGGCT AGTTTCCTAG TTTGATCTTT GATTTTCATT GAGTATTAGT	2700
AATTCAGTTA CTAATCGTC AACTCTGATT TATCCAATAA AATTGAAAAG GATGGAAAAA	2760
AGGATAAATT TATGATATAC TTTATTTTGA AGACCTTATT AGAAATCTTG AAAGAGTATT	2820
GAAAACTTAG AATGAGAAAA ATTGTTATCA ATGGTGGATT ACCACTGCAA GGTGAAATCA	2880
CTATTAGTGG TGCTAAAAAT AGTGTGCTTG CCTTAATTC AGCTATTATC TTGGCTGATG	2940
ATGTGGTGAC TTTGGATTGC GTTCCAGATA TTTCGGATGT AGCCAGTCTT GTCGAAATCA	3000
TGGAATTGAT GGGAGCTACT GTTAAGCGTT ATGACGATGT ATTGGAGATT GACCCAAGAG	3060
GTGTTCAAAA TATCCAATG CCTTATGGTA AAATTAACAG TCTTCGTGCA TCTTACTATT	3120
TTTATGGGAG CCTCTTAGGC CGTTTGGTG AAGCGACAGT TGGTCTACCG GGAGGATGTG	3180
ATCTTGGTCC TCGTCCGATT GACTTACACC TTAAGGCGTT TGAAGCTATG GGTGCCACTG	3240
CTAGCTACGA GGGAGATAAC ATGAAGTTAT CTGCTAAAGA TACAGGACTT CATGGTGCAA	3300
GTATTTACAT GGATACGGTT AGTGTGGGAG CAACGATTAA TACGATGATT GCTGCGGTTA	3360
AAGCAAATGG TCGTACTATT ATTGAAAATG CAGCCCGTGA ACCTGAGATT ATTGATGTAG	3420

185

CTACTCTCTT GAATAATATG GGTGCCCCATA TCCGTGGGGC AGGAACTAAT ATCATCATTA	3480
TTGATGGTGT TGAAAGATTA CATGGGACAC GTCATCAGGT GATTCCAGAC CGCATTGAAG	3540
CTGGAACATA TATATCTTTA GCTGCTGCAG TTGGTAAAGG AATTCGTATA AATAATGTTT	3600
TTTACGAACA CCTGGAAGGG TTTATTGCTA AGTTGGAAGA AATGGGAGTG AGAATGACTG	3660
TATCTGAAGA CAGCATTTTT GTCGAGGAAC AGTCTAATTT GAAAGCAATC AATATTAAGA	3720
CAGCTCCTTA CCCAGGCTTT GCAACTGATT TGCAACAACC GCTTACCCCT CTTTACTAA	3780
GAGCGAATGG TCGTGGTACA ATTGTCGATA CGATTACGA AAAACGTGTA AATCATGTTT	3840
TTGAACTAGC AAAGATGGAT GCGGATATTT CGACAACAAA TGGTCATATT TTGTACACGG	3900
GTGGACGTGA TTTACGTGGG GCCAGTGTTA AAGCGACCGA CTTAAGAGCT GGGGCTGCAC	3960
TAGTCATTGC TGGGCTTATG GCTGAAGGTA AACTGAAAT TACCAATATC GAGTTTATCT	4020
TACGTGGTTA TTCTGATATT ATCGAAAAAT TACGTAATTT AGGAGCGGAT ATTAGACTTG	4080
TTGAGGATTA AACCGTAGAG GTGTTTATGA ATATTGGAC CAAATTAGCA ATGTTTTCTT	4140
TTTTTGAAAC GGATCGCTTG TATTTGCGTC CTTTCTTTTT TAGTGATAGT CAGGACTTCC	4200
GCGAGATAGC TTCAAATCCA GAAAATCTTC AATTTATTTT CCCAACGCAG GCAAGTCTGG	4260
AAGAAAGTCA ATATGCACTG GCCAATTACT TTATGAAGTC CCCTTTGGGA GTGTGGGCAA	4320
TTTGTGACCA GAAAAATCAA CAAATGATTG GTTCTATTAA ATTTGAGAAG TTAGATGAAA	4380
TCAAAAAGA AGCTGAGCTT GGCTATTTTT TGAGAAAAGA TGCTTGGTCG CAAGGATTTA	4440
TGACAGAGGT TGTTAGAAAA ATTTGTCAGC TTTCTTTTGA GGAATTGGC TTAACAAT	4500
TATTTATCAT TACCCACCTT GAAAATAAAG CTAGCCAAAG AGTTGCTCTT AAGTCTGGAT	4560
TTAGTTGTT CCGTCAGTTT AAGGAAGTG ATCGTTACAC AAGAAAAATG CGGGATTATC	4620
TTGAATTTTCG GTATGTAAAA GGAGAGTTCA ATGAGTAAGC ATCAGGAAAT TCTAAGCTAT	4680
TTGGAGGAAT TACCAGTAGG TAAAAGGGTC AGTGTTCGTA GCATTTCGAA TCATCTAGGA	4740
GTTAGTGATG GAACAGCCTA TCGGGCTATT AAAGAAGCTG AAAACCGTGG AATTGTGGAG	4800
ACCCGTCCTA GAAGTGAAC AATTCGTGTT AAATCCAGA AAGTTGCTAT AGAGAGATTA	4860
ACGTTTGCTG AAATTGCAGA AGTGACTTCT TCTGAGGTTT TGGCTGGGCA AGAAGGTTTA	4920
GAGAGAGAAT TTAGTAAGTT TTCAATTGGT GCCATGACTG AACAAAATAT CTGTCTTAC	4980
CTTCATGATG GGGGGCTCTT GATTGTCGGA GACCGAACCC GTATTCAGTT GCTAGCCTTG	5040
GAAAATGAAA ATGCACTTCT GGTTACAGGG GGATTTTCAGG TTCATGATGA TGTGCTTAAA	5100
CTGGCCAATC AAAAAGGGAT TCCTGTTCTA AGAAGTAAGC ATGATACCTT TACCGTCGCG	5160

186

ACCATGATCA ATAAAGCCTT GTCAAATGTC CAAATCAAGA CTGATATTCT GACAGTTGAG	5220
AAACTTTATC GCCCTAGTCA TGAGTATGGT TTTCTGAGAG AGACAGATAC AGTTAAAGAT	5280
TATTTGGACT TGGTTCGTAA GAATCGTAGC AGCCGTTTCC CTGTTATCAA TCAACATCAG	5340
GTCTGTGTTG GTGTTGTAAC CATGAGAGAC GCTGGTGATA AATCACCAAG CACGACAATT	5400
GATAAGGTTA TGTCTCGTAG TCTATTTTGG GTTGGATTAT CGACAAATAT TGCCAATGTG	5460
AGTCAACGGA TGATCGCAGA AGACTTTGAA ATGGTACCAG TTGTTGGAAG CAATCAAAC	5520
TTGCTTGCGC TTGTGACGCG ACGAGATGTC ATGGAGAAGA TGAGCCGTTT CCAAGTTTCG	5580
GCTCTACCAA CTTTCTCTGA GCAGATTGGA CAAAAGCTCT CTTATCACCA TGATGAAGTA	5640
GTCAATACAG TGGAAACCTT TATGCTAGAA AAAAATGGAG TTTTGGCTAA TGGTGTATTG	5700
GCAGAAATTC TGACCCACAT GACCCGATTT AGTTGTTAAT AGTGGTCGCA ATCTCATTAT	5760
CGAGCAGATG CTGATCTACT TTTTGCAGGC TGTTCAGATA GATGATATAT TGCGCATTCA	5820
GGCACGGATT ATTCATCATA CGAGACGGTC AGCTATAATT GATTACGATA TTTATCATGG	5880
TCACCAGATT GTTTCAAAAG CAAATGTGAC TGTTAAAATT AATTAGAAAC TAGGAGAAAA	5940
GATGATAACA TTAATATCAG CTCGTGAAAT CGAAGCTATG GACAAGGCTG GTGATTTTCT	6000
AGCAAGTATT CATATAGGCT TACGTGATTT GATTAAGCCA GGCGTAGATA TGTGGGAAGT	6060
TGAAGAATAT GTCGCGCGTC GTTGTAAGA AGAAAATTTT CTTCCACTTC AGATTGGGGT	6120
TGACGGTGCC ATGATGGACT ATCCTTATGC TACCTGTTGC TCTCTTAACG ATGAAGTGGC	6180
TCACGCTTTC CCTCGTCATT ATATCTTGAA AGATGGTGAT TTGCTCAAAG TTGATATGGT	6240
TTTGGGAGGT CCCATTGCTA AATCTGACCT AAATGTCTCA AAATTAAC TCAACAATGT	6300
TGAACAAATG AAAAAATACA CTCAGAGCTA TTCTGGTGGT TTAGCAGACT CATGTTGGGC	6360
TTATGCTGTT GGTACACCGT CCGAAGAAGT CAAAACTTG ATGGATGTAA CCAAAGAAGC	6420
TATGTACAAG GGTATTGAGC AAGCTGTTGT TGGAAATCGT ATCGGTGATA TCGGTGCGGC	6480
TATTCAAGAA TACGCTGAAA GTCGTGGTTA CGGTGTAGTG CGTGATTTGG TTGGTCATGG	6540
TGTTGGCCCA ACTATGCACG AAGAACCAAT GGTTCCTAAC TATGGTATTG CAGGTCGTGG	6600
ACTCCGTCTT CGTGAAGGAA TGGTCTTAAC CATTGAACCA ATGATCAATA CAGGCGATTG	6660
GGAAATTGAT ACAGATATGA AAAGTGGTTG GCGCATAAG ACCATTGACG GTGGATTGTC	6720
ATGTCAGTAT GAACACCAAT TTGTCATTAC GAAAGATGGA CCTGTTATCT TGACTAGCCA	6780
AGGTGAAGAA GGAACCTTAT AATAAAAAGT GAAAAGACTA CTGGAAGTTT ATTTTGATAA	6840
AAAATCCAGT AGATCTTTTC ATAATAAAC GCATTGTATC AAGTGTAGG GGCTGATATC	6900
ATGCGTTTTT CTGCTTTTAA GATTTTTTCC AACTCTGTTT GTAAGCGCAT CATAACAAAG	6960

GGTCTAGGAT TCAGGGCTCT CCTCCTATAT ACTATTAGTA AAGTAAACT AAGGGAGGAT	7020
ATTTTAGTGT CGCAGTCTAT TGTTCCTGTA GAGATTCCAC AATATTGTCG TTTTGATTCT	7080
AAAAGAGAA ATGGAATTCT GTTAAATGTT CGTATTGCCA ATCTTAAAT TACTTTTTTA	7140
TATTATACTT CCTGCGAAAC AAAATATGGT ATAGTAGTTC TATGAATGAT GAAGCAAGTA	7200
AACAATAAC TGATGCACGA TTTAAGCGTC TTGTTGGTGT TCAGCGTACC ACTTTTGAAG	7260
AGATGTTAGC TGTATTAAAA ACAGCTTATC AACTTAAACA CGCAAAAGGT GGACGAAAAC	7320
CTAAATTAAG CCTAGAAGAC CTTCTTATGC CCACTCTTCA ATAGTCCGAG AATATCGAAC	7380
TTATGAAGAA ATTGCGGCTG ATTTTGGTAT TCACGAAAGC AACTTTATCC GTCGGAGCCA	7440
ATGGGTTGAA ATAACCTCTG TTCAAAGTGG TTTTACGGTT TCAAGAACTC CTCTCAGTTC	7500
TGAGGACACG GTAATGATTG ATGCGACGGA AGTAAAAATC AATCGCCCTA AAAAAACAAT	7560
TAGCGAATGA TTCTGGTAAA AAGAAATTC ACCTATGAA GGCTCAAGCG ATTGTCACAA	7620
GTCAAGGGAG AATTGTTTCT TTGGATATCG CTGTGAACTA TAGTCATGAT ATGAAGTTGT	7680
TCAAATGAG TCCTAGAAAT ATCGAACAAG CTGGTAAAT CTTGGCTGAC AGTGTTATC	7740
AAGGGCTCAT GAAGATATAT CCTCAAGCAC AACTCCACG TAAATCCAGC AACTCAAGC	7800
CGCTAACAGC TGAAGATAAA GCCTATAACC ATGCGCTATC TAAGGAAAGA AGCAAGGTTG	7860
AGAATCTTT TGCCAAAGTA AAAACGTTTA AAATATTTT AACAACCTAT CGAAATCATC	7920
GTAAACGCTT CGGATTACGA ATGAATTTGA GTGCTGGTAT TATCAATCAT GAACTAGGAT	7980
TCTAGTTTGT CAGGAAGTCT ATTGAGGTAT TGAGCTAGTT TATGAAAAA TTGGGTGAAA	8040
AGTCGAGTGT TTTAGAAACC CACAGTGTAG TATTCTAGTT TCAATCCACT ATATTTTGCT	8100
ACTCCCGTA AAGTTTCTAT TTTCCCTGAT TTCTGATATA ATAGAAATAT TGAATCAAG	8160
AGTAAGGAAG AGAAGATGAA CGCATTATTA AATGGAATGA ATGACCGTCA GGCTGAGCG	8220
GTGCAAACGA CAGAAGGTCC CTTGCTAATC ATGGCAGGGG CTGGTTCTGG AAAGACTCGT	8280
GTTTGTGACCC ACCGTATCGC TTATTTGATT GATGAAAAGC TGGTCAATCC TTGGAATATC	8340
TTGGCCATTA CCTTTACCA CAAGGCTGCG CGTGAGATGA AAGAGCGTGC TTATAGCCTC	8400
AATCCAGCGA CTCAGGACTG TCTGATTGCG ACCTTCCACT CCATGTGTGT GCGTATTTTG	8460
CGTCGCGATG CGGACCATAT TGGCTACAAT CGTAATTTTA CAATTGTGGA TCCTGGTGAA	8520
CAGCGAACGC TCATGAAACG TATTCTCAA CAGTTGAACT TGGACCTAA AAAATGGAAT	8580
GAACGAACTA TTTTGGGGAC CATTTCCAAT GCTAAGAATG ATTTGATTGA TGATGTTGCT	8640
TATGCTGCCC AAGCTGGCGA TATGTATACG CAAATTGTGG CCCAGTGTTA TACAGCCTAT	8700

188		
CAAAAAGAAC TTCGTCAGTC TGAATCCGTT GACTTTTGATG ATTTGATTAT GCTGACCTTG	8760	
CGTCTCTTTG ATCAAAATCC TGATGTTTTG ACCTACTACC AGCAAAAATT CCAATACATC	8820	
CACGTTGATG AGTACCAAGA TACCAACCAC GCTCAGTACC AATTGGTCAA ACTCTTGGCT	8880	
TCCCGTTTTA AAAATATCTG TGTGGTTGGG GATGCGGACC AGTCTATCTA CGGTTGGCGT	8940	
GGTGCTGATA TGCAGAATAT CTTGGACTTT GAAAAGGATT ACCCCAAAGC CAAGGTTGTT	9000	
TTGTTGGAGG AAAATTACCG CTCAACCAA ACCATTCTCC AAGCGGCCAA CGAGGTTATT	9060	
AAAAATAATA AAAATCGCCG TCCTAAAAAT CTCTGGACTC AAAACGCTGA TGGGGAGCAA	9120	
ATCGTTTACT ATCGTGCCGA TGATGAGCTG GATGAGGCTG TATTGTAGC CAGAACCATC	9180	
GATGAACTTA GTCGAGTCA AAACCTCCTT CATAAGGATT TTGAGTTCT CTATCGGACT	9240	
AATGCCCAGT CCCGTACAAT TGAGGAAGCC CTGCTCAAGT CTAACATTCC TTATACCATG	9300	
GTTGGCGGAA CCAAATTCTA CAGCCGTAAG GAAATTCGCG ATATTATTGC TTATCTCAAC	9360	
CTTATTGCTA ATTTGAGTGA CAATATTAGT TTTGAGCGTA TTATCAACGA GCCTAAACGT	9420	
GGAATTGGTC TAGGTACAGT TGAGAAAATC CGTGATTTTG CAAATTTGCA AAATATGTCT	9480	
ATGCTGGATG CTTCTGCTAA TATTATGTTG TCTGGTATCA AGGGTAAGGC AGCCCAATCT	9540	
ATCTGGGATT TTGCCAATAT GATGCTTGAT TTGCGGGAGC AGCTAGACCA CTTAAGCATT	9600	
ACAGAGTTGG TTGAGTCCGT CCTAGAAAA ACAGGTTATG TCGATATTCT TAACTCCCAA	9660	
GCGACTCTAG AAAGCAAGGC ACGGGTTGAA AATATCGAAG AGTTTCTTTC TGTTACGAAG	9720	
AACTTTGATG ACACCACGGA TGTGACAGAA GAGGAAACTG GTCTGGACAA ACTGAGTCGT	9780	
TTCTTAAATG ACTTGGCTTT GATTGCCGAC ACAGATTCAG GTAGTCAGGA GACATCAGAA	9840	
GTGACCTGA TGACCCTGCA TGCTGCCAAA GGTCTCGAAT TTCCAGTTGT CTTTTTGATT	9900	
GGGATGGAAG AAAATGTCTT TCCACTTAGT CGTGCGACTG AAGATTCAGA TGAATTAGAA	9960	
GAAGAGCGCC GTCTAGCCTA TGTAGGTATC ACGCGTGACG AGAAAATTCT CTATCTGACC	10020	
AATGCCAACT CAGCTTGCT TTTTGGTTCG ACCAATTATA ACCGTCCGAC TCGTTTTATT	10080	
AACGAAATCA GTTCAGACTT GCTTGAGTAT CAAGGTCTGG CTCGTCCTGC AAATACAAGC	10140	
TTTAAGGCAT CATATAGCAG TGGTAGTATT TCCTTTGGTC AAGGTATGAG TTTGGCTCAG	10200	
GCTCTTCAAG ACCGTAAACG CGGTGCTGCC CCAAAATCAA TCCAGTCAAG CGGTCTTCCA	10260	
TTTGGTCAAT TTACAGCTGG CGCAAAACCA GCATCTAGCG AGGCAAATTG GTCCATTGGT	10320	
GATATTGCTC TCCACAAGAA ATGGGGAGAG GGAACCGTTC TGGAAGTTTC AGGTAGCGGT	10380	
GCTAGGCAGG AATTGAAAAT CAATTTCCCA GAAGTAGGTT TGAAAAAACT TTTAGCCAGT	10440	
GTGGCTCCAA TTGAGAAAAA AATCTAATTT TCCATCCTTC TCACGAATAA TAAAGTGAGG	10500	

AGGATTTTTA TGTACAGTAT TTCATTCCAA GAAGATTCAC TATTACCAAG AGAAAGGCTG	10560
GCCAAGGAAG GAGTTGAAGC GCTTAGTAAC CAAGAGTTGC TAGCTATTTT ACTCAGGACA	10620
GGAACACGTC AAGCTAGCGT TTTTGAAATT GCCCAAAAAG TCTTGAACAA TCTTTCAAGC	10680
CTAACGGATT TGAAAAAAT GACCCTGCAG GAATTGCAGA GTTGTCTGG TATTGGGCGT	10740
GTTAAGGCCA TAGAATTACA AGCTATGATT GAACTGGGGC ATCGTATTCA CAAACACGAG	10800
ACTCTTGAAA TGGAAGTAT TCTCAGCAGT CAAAAGTTGG CCAAGAAGAT GCAGCAGGAA	10860
TTAGGGGATA AAAACAAGA GCACCTGGTG GCACTCTATC TCAATACTCA AAATCAAATC	10920
ATCCATCAGC AGACCATTTT TATCGGGTCT GTAACCTCGTA GTATCGCTGA ACCGCGAGAG	10980
ATTCTTCACT ATGCAATCAA GCATATGGCG ACTTCTCTTA TCTTGGTCCA CAATCATCCT	11040
TCAGGAGCGG TAGCGCCTAG CAAAATGAT GATCATGTCA CTAAACTTGT TAAAGAAGCC	11100
TGCGAATTGA TGGGGATTGT TCTCTTGGAC CATTTGATTG TCTCTCATTG TAATTACTTT	11160
AGTTATCGTG AAAAGACAGA TTTAATCTAA AGTTCATTAA CGACATAGTC AAAGAGTTTT	11220
TTATCTTTGG GACGATTTTC AAAAAGAAGT TCTGGATGCC ATTGGACACC GAGAAAGGCG	11280
ACATCATCCG TACTCATGAC AGCCTCAATG ATACCATCTT TAGGATCATG AGCCACAAC	11340
TTTAAATTG GTGCTAAGTC CTTGATGCTC TGGTGGTGA AGGAGTTGAT ATGAGAGATT	11400
TCTCCATAGA TTTCTTGGAG AACGGTATCT GGTCTGTGA CCAAGCGTTG AGTTGTGTAC	11460
TCAACAGAAG AATCCTGCCA ATGGTCTCG ATATCTTGGT ACAAAGTTCC ACCCATGGCA	11520
ACGTTAAAGA GTTGGGTACC ACGGCAGACA GAGAAAATGG GCTTTTCTG TTTAATAGCT	11580
TCCTTGATGA GGGCCAGTTC GAAGATATCT CTTTGAAGGT GATAGTCATC ACTATCAATG	11640
GTTTTGGGTT CGCCATAAAA TTTTGGATCG ACATTTTGCC CACCTGTCAA GATGAGCTTG	11700
TCAATCAAAC TGATATAGTG GCAGGCCATT TCTTGATCAC CAATCGGTAG GATGATGGGA	11760
ATCCCTCCAG CATCTTTAAC GCCTTCAACA AAGCCTTTTG CTGCGTAGCT CATCATGATG	11820
TCATCATCTG GATGAGTTTT TTCGTTTCCT GTAATCCCAA TAACTGGTTT TTTCATAAAA	11880
TGATTTTCGC TTTCTAATCC TCTTTTCGCA TGAAGTAGAG GAGGGTTTGG AGTTCACCTG	11940
TCAAATCGAC ATACTGAACG ACCACGTCTT TTGGTAAATG CAGATGGACT GGTGAAAAAC	12000
TGAGAAATCC TTTACACCA GCATCAACCA AGAGATTAGC AACCTCTTGT GACTTGACGC	12060
TGGGAACAGT TAGGATAGCA GTCTTCACAT CAGCATCCTT GATTTTATCC TTGATCTGAG	12120
AAATCCCGTA AATGGGAATC CCGTCAGGAG TTTGGGTACC GACTTCAGGA TGGTCGTCTA	12180
GGTCAAAGGC CATGATAATC TTCATCTTGT TACGTTCTGT GAAGCGGTAG TGGAGAAGGG	12240

190

CATGGCCCAT ATTTCCAATA CCAACCAGCA TGACATTGGT AATAGAGTTG TCATTGAGCA	12300
AATCGGCAAA AAATGTCATT AGTTTMTTGA CATCATAGCC AAAACCACGA CGACCAAGTT	12360
CACCAAAATA GGAAAAATCA CGACGTACGG TCGCTGAATC AATACCGATA GCCTCTGCAA	12420
TTTGCTTAGA GTTGGCACGT TCAATCTTTT CTGCATGAAA TCTCTTAAAA ATTCGATAGT	12480
AGAGAGAGAG TCTTTTGTCT GTAGCTTTTG GAATAGCAAA CTGTTTATCT TTCACAAAAT	12540
CACAACCTTT CTATTCTTCT ATTTTATAGA AACATTGTGA AAAATCAAC AAAATAAGA	12600
AAAACTAAG AAAATCTTA GTTTTGATGT AAAAACTCTG CATGAGATAG AAAACGGTAG	12660
AGGTCTCCGA CCAGCCCCTG ATAACTTTT TTGCCCCTAA AAGTCAGAGA AGTCACATAA	12720
ACTGTATCTG GTAAGGTAC ACATCCTGAC AAAGTCAACA TGAGAGCCTC ATGATCCTCA	12780
TACTTGAGAG TACGCTCTAC ATGATAGCAG TCCTTATAGG TCAGTTCAA CATTMTGGCT	12840
CTATCTTTCC GATTTGTAA AGACACCAG TTCTACCAAG CTATCCATGA GGAAGTAGAA	12900
TTTTCTCTGA TGAATATGGT GGTCTTCTGA TTGAAAATA TCAACTAGAC GAAGGCCAAA	12960
CTTGTCAGTG ATATTGATTT TAGCCCCTGT AAGTTCCTTG TTAATGATGA TTTTGAGTTG	13020
GAAGCCTTCA CCGCTGTTTG GCACTTTTTC CAAAAGGCGA GTCAGTTCAT AGTTACCAAC	13080
CTTAGTTTCA AAAAAGGTGT TATCTTGAG GGTGAATTTT TTAACAGAAG GGCTAAGAGT	13140
GTAATCGTAA CGACAATTTT TTAAGTGAAT GATTTTTTCA AATGCCATAT GGCTAACCTC	13200
CGATAATTTT TTTTAAGGTT TTTGCGAGGG TTTGTAGGTC TTCAACGGTA TTTTGTGGCG	13260
ACAACTGAT GCGAAGGGAT TCCTTCAAGC GTTCTGAATT TGCGCCATAC ATGGCTTCAA	13320
GAACATGGCT GGATTGGACA ACGCCTGCAG TACAGGCTGA GCCAGTAGAG ATTGAAATTC	13380
CAGCTAAATC TAGCCGAAGG AGTAAGAGGT CATTTTCTG ACCAGGAAAT CCAATATTGA	13440
GAACATAAGG GAGATGATGT TTTCTCTAT TCAGGTAATA CTGAATGCCC TCCAGCTCTG	13500
CCAGAAAGGC AGTTTCTAGA TTTGTACAT GTTGAAATG TTCTTCTGT TTTTCTAGGT	13560
CTTCTTTTAG GGCTGCAACC ATGCCTACAA TGGCAGGCAG ATTTTCAGTT CCTGCACGTT	13620
TTTTCTGTTC CTGGTCTCCG CCATGTAGAT AGGAATCAAA GTCCATGCTA GATGCGTAGA	13680
GAAAACCGAT TCCCTTAGGA CCATGGAATT TGTGGGCAGA AGCAGTGAGA AAATCAATGC	13740
CCAATTCTTC TGAATGAATT GGGATTTTAC CAATAGCCTG AACTGCATCA ACATGATAGG	13800
CAGCAGGGTG TTGCTTGAGT ATTTGGCCAA TTTCAGCGAT GGGCAGTAGG TTTCTGTCT	13860
CATTATTGAC AAACATGGTA GAAACCAAAA TCGTATCGTC ACGTAAAGCC TTTTGAATTT	13920
GCTGGGCTGT GATTTCTTGA TTTTCTGGCT GGATAATGGT TGCTTCAAAC CCAAAGTGT	13980
GAACCAAGTA ATCAATTGTT TCAAGGACAG CATGGTGCTC GATGGCAGTT GTGATGATAT	14040

GTTCCTTC	TTCTTGGTGA	CGAAGACAGT	AGCCAATGAT	GGTAGTATTA	TTGCCTTCAG	14100
TCCCACCAGA	AGTGAAAAAG	ATATGTTGAG	GTTTGTGCCT	TAGTAACTGG	GCTAGTTCCT	14160
GACGGGCTTC	TCGCAAGAGT	TTGCCAGCTT	GACGACCATG	ACCATGAATA	CTAGAAGGAT	14220
TTCCGTGGGT	TTCTTGCAATA	ACCTTGGTCA	TAGCTGAAAT	AGCAACTGCT	GACATAGGAG	14280
TCGTTGCAGC	ATTGTCCAAA	TAAATCAAAG	AATCACCTTA	TTTCTTTTTA	TTGTAGGCAA	14340
AGAGTGGGCT	GACTGGTTTT	CTTTCGTGAA	TACGGACGAT	AGCATCACCA	ATTAATCTAC	14400
TAGCAGTGAT	GTAGCATACA	TTTTTAGGAG	TTTTTTCITT	TGTTGCTACT	GAATCAGTCA	14460
CAAGAATTTC	TTTAATATTA	GTATTGTCAA	GAAGCTCAGC	AGCTCCCTCG	ACGAAGAGAC	14520
CGTGGCTAGA	AACAGCATAA	ATTTCTGTAG	CTCCTTCACG	TTCAACGATT	TTAGAAGCTT	14580
CAGAGAAGGT	ACGTCCTGTA	TTTAAATAT	CATCAATCAA	GATAGCTTTC	TTACCTTCAA	14640
CATCACCAAT	AATATAACCT	TCGTTACGAG	TTGCATCGTC	TTGAGGGTAG	TCGATAATGG	14700
CGATAGGAGC	ATCAAGATAT	TCAGCCAGGC	TACGCGCACG	TTTGACACCT	GAATTTTTAG	14760
GGCTAACGAC	AACAACATCT	GAACCAAGCA	ATCCTTTATC	GCAGTAATGT	TTTGCGAATA	14820
GGGGAACAGT	GAAAAGATTA	TCCACTGGAA	TATCAAAGAA	ACCTTGAACC	TGAACGGCAT	14880
GCAAAATCAAG	AGTCAGGATA	CGATCAACTC	CAGCCTTAAC	CAGCATATTG	GCAACTAGTT	14940
TTGCTGTAAG	TGGCTCACGA	GGACAAGCAA	TGCGGTCTTG	ACGTGCATAG	CCAAAATATG	15000
GAAGGACAAC	GTTGATACTG	TGGGCACTTG	CACGCACACA	AGCATCGACC	ATGATTAACA	15060
ATTCATTAG	GTGGTTGTTG	ACAGGGAAAC	TTGTTGATTG	GATGATGTAA	ACATCATAAC	15120
CACGGACACT	TTCTTCGATA	TTTACTTGGA	TTTCTCCGTC	TGAAAATGA	CGTGATGATA	15180
GTTTCCAAG	TGGGACACCA	ACAGCTTGGG	CAATTTTTTG	TGCAATCTCT	TGGTTAGAGT	15240
TGAGTGCGAA	AAGTTTCATG	TTTTTCTAT	CTGACATTAT	AGACCGTCTT	CTGTAAACTT	15300
TATAAATCCT	AGTTATATTT	ACCTTACATA	TATGAACTGG	GATTTGTGTA	TTTTTATCTT	15360
TTCTATTTTA	CCAAAAAATG	GAGATTATTT	CAGCTATTTT	TCATACTTTT	GACAAATCGA	15420
ACCAATTTTG	AAGGAGCTTT	TTGATAGGAA	ATCTGATTTT	TCTCTAAAAA	TTGTCGAAAA	15480
TCCTGTTTGC	CTTGCTCATG	ATTTTCCACT	TCAAGCTCCA	ATTCGTAATC	TGTTATATCA	15540
AAGTATCGGC	TCTGATCCAG	TGCCATGAGA	CCAATAGCTG	TTTTCATTTT	ATAGCGAAGC	15600
GTTGTTAGAC	AACCAAGAAC	CTGCCAGTTC	TTACTTTGGA	TACCATGTTT	CGCCAATTCA	15660
TCCAGTACTA	GCCCTTGAGG	AAGTCTTCC	TTACTCAGAT	AGTCTCAGC	ATCTTTTAGT	15720
TGCAATTTTT	GGTTGTATTC	CATGTTTCCA	ACACTCTGCG	GGACTTTGAG	TGTCAACTCA	15780

192

GCCCAGTCTT CAAAGGTCG AATGCGCATA GCGACTTTCT TTTCTCGCAG TTCAAAATCA	15840
GGCGTGTCGA TGTAGTAATT TGTTTGAAGA ACAGGAGTGA CACCTGTGAA CTGGTCTTTT	15900
AGACGATTGT ATTCATCTTT TTTCAATAGT GTTTCAATT CAATTTCTAA ATGTTTCATT	15960
TTTCTTACCT TTTTTTATCG TTGAAAGCGG ATTTATGGTA TAATAAGCAT TGTATTTATT	16020
GTATATGAAT CTGGAGAAAA AATCAAAGAT ATTTTGTGACG GATAATATGA GAACAAGGGA	16080
GAATATATGA CCTAGAAATG GGAAGAATTT CTAGATCCTT ACATTCAACC TGTGGTGAG	16140
TTAAAGATTA AACTTCGTGG TATTCGTAAG CAATATCGTA AGCAAAATAA GCATCTCCA	16200
ATTGAGTTTG TGACCGGTCG AGTCAAGCCA ATTGAGAGCA TCAAAGAAAA AATGGCTCGT	16260
CGTGGCATT A CTATGCGAC CTTGGAACAC GATTTGCAGG ATATTGCTGG CTTACGTGTG	16320
ATGGTTCAGT TTGTAGATGA CGTCAAGGAA GTAGTGGATA TTTTGCACAA GCGTCAGGAT	16380
ATGCGAATCA TACAGGAGCG AGATTACATT ACTCATAGAA AAGCATCAGG CTATCGTTCC	16440
TATCATGTGG TAGTAGAATA TACGGTTGAT ACCATCAATG GAGCTAAGAC TATTTTGGCA	16500
GAAATTCAAA TTCGTACTTT GGCCATGAAT TTCTGGGCAA CGATAGAACA TTCTCTCAAC	16560
TACAAGTACC AAGGGGATTT CCCAGATGAG ATTAAGAAGC GACTGGAAAT TACAGCTAGA	16620
ATCGCCCATC AGTTGGATGA AGAAATGGGT GAAATTCGTG ATGATATCCA AGAAGCCCAG	16680
GCACTTTTTG ATCCTTTGAG TAGAAAAATTA AATGACGGTG TAGGAAACAG TGACGATACA	16740
GATGAAGAAT ACAGGTAAAC GAATTGATCT GATAGCCAAT AGAAAACCGC AGACTCAAAG	16800
GGTTTTGTAT GAATTGCGAG ATCGTTTGAA GAGAAATCAG TTTATACTCA ATGATACCAA	16860
TCCGGATATT GTCATTTCCA TTGGCGGGGA TGGTATGCTC TTGTCGSCCT TTCATAAGTA	16920
CGAAAATCAG CTTGACAAGG TCCGCTTTAT CGGTCTTCAT ACTGGACATT TGGGCTTCTA	16980
TACAGATTAT CGTGATTTTG AGTTGGACAA GCTAGTGAAT AATTGTCAGC TAGATACTGG	17040
GGCAAGGGTT TCTTACCCTG TTCTGAATGT GAAGGTCTTT CTTGAAAATG GTGAAGTTAA	17100
GATTTTCAGA GCACTCAACG AAGCCAGCAT CCGCAGGTCT GATCGAACCA TGGTGGCAGA	17160
TATTGTAATA AATGGTGTTC CCTTTGAACG TTTTCGTGGA GACGGGCTAA CAGTTTCGAC	17220
ACCGACTGGT AGTACTGCCT ATAACAAGTC TCTTGGCGGT GCTGTTTTAC ACCCTACCAT	17280
TGAAGCTTTG CAATTAACGG AAATTGCCAG CCTTAATAAT CGTGTCTATC GAACACTGGG	17340
CTCTTCCATT ATGTGCCTA AGAAGGATAA GATTGAACTT ATTCCAACAA GAAACGATTA	17400
TCATACTATT TCGGTTGACA ATAGCGTTTA TTCTTTCCGT AATATTGAGC GTATTGAGTA	17460
TCAAATCGAC CATCATAAGA TTCACTTTGT CGCGACTCCT AGCCATACCA GTTCTGGAA	17520
CCGTGTTAAG GACGCCTTTA TCGGCGAGGT GGATGAATGA GGTTTGAATT TATCGCAGAT	17580

GAACATGTCA AGGTTAAGAC CTTCTTAAAA AAGCACCAGG TTTCTAAGGG ATTGCTGGCC	17640
AAGATTAAGT TTCGAGGTGG AGCTATTCTG GTCAATAATC AACCGCAAAA TGCAACGTAT	17700
CTATTGGACG TTGGAGACTA CGTTACCATT GACATTCCCG CTGAGAAAGG CTTTGAAACC	17760
TTGGAGGCTA TTGAGCTTCC ATTAGATATT CTCTATGAGG ATGACCACTT TCTAGTCTTG	17820
AATAAACCCCT ATGGAGTGGC TTCTATTCCCT AGTGTCAATC ACTCTAATAC CATTGCCAAT	17880
TTTATCAAGG GTTACTATGT CAAGCAAAAT TATGAAAATC AGCAGGTTCA CATTGTTACC	17940
AGACTAGATA GGGATACTTC TGGCTTGATG CTCTTTGCCA AGCACGGTTA TGCCCATGCA	18000
CGATTAGACA AGCAGTTGCA GAAGAAATCT ATCGAGAAAC GCTACTTTGC TTTGGTTAAG	18060
GGAGATGGAC ATTTGGAGCC AGAAGGGGAA ATTATTGCTC CGATTGCGCG TGATGAAGAT	18120
TCCATTATTA CCAGACGAGT GGCTAAAGGC GGAAAGTATG CCCATACTTC ATACAAGATT	18180
GTAGCTTCTT ATGGAATAT TCACTTGCTC TATATTCACC TGCACACTGG TCGAACCCAT	18240
CAAATCCGAG TCCATTTTTC TCATATCGGT TTTCTTTGC TGGGACATGA TTTGTATGGT	18300
GGTAGTCTGG AAGATGGTAT TCAACGTCAG GCTCTGCATT GCCATTACCT ATCCTTTTAT	18360
CATCCATTTT TAGAGCAAGA CTGTCAGTTA GAAAGTCCCT TGCCGGATGA TTTTAGTAAC	18420
CTTATTACCC AGTTATCAAC TAATACTCTA TAAAACTGT CTCAGAGTAT AATTATTATC	18480
TTAAAGGAGA AAACATCATG AAGTTTGTGA AAGTCTCAA GCCAACCTTG TTGGTAAAAA	18540
TGCTCGTATC GTTCTCCCTG AAGGGGAAGA GCCTCGTATT CTTCAAGCAA CAAAACGCTT	18600
AGTAAAAGAA ACAGAAGTGA TTCCTGTTTT GCTTGGAAT CCTGAAAAA TTAAAATTTA	18660
TCTTGAAATT GAAGGAATCA TGGATGGTTA TGAGGTCATC GACCCCTAAC ATTATCCTCA	18720
ATTGAAGAA ATGGTTTCTG CCTTGGTGGA GCGTCGCAAG GGCAAAATGA CTGAAGAAGA	18780
TGTACGCAAG GTTTTGGTTG AAGATGTCAA CTACTTTGGT GTGATGTTGG TTTACTTGGG	18840
CTTGGTTGAT GGAATGGTGT CAGGAGCGAT TCACTCAACA GCTTCAACAG TTCGCCCAGC	18900
TCTACAAATC ATCAAACTC GTCCAAATGT AACTCGTACT TCAGGAGCCT TCCTCATGGT	18960
TCGTGGTACG GAACGTTACC TATTTGGAGA CTGTGCCATT AACATCAATC CAGATGCAGA	19020
AGCCTTGGCT GAAATTGCCA TCAACTCAGC AATCACAGCT AAGATGTTTG GCATCGAACC	19080
TAAAATTGCC ATGTTGAGCT ATTCTACTAA AGGTTGAGG TTTGGTGAAA GCGTTGATAA	19140
GGTCGTTGAA GCAACTAAAA TTGCTCACGA CTGCGTCCT GACCTTGAAA TCGATGGTGA	19200
GTTGCAATTT GATGCAGCCT TTGTTCTGA AACTGCAGCT CTGAAAGCTC CTGGAAGTAC	19260
GGTAGCTGGT CAAGCAAATG TCTTCATCTT CCCAGGTATC GAGGCAGGAA ATATTGGTTA	19320

194

CAAGATGGCT GAACGCCTGG GTGGCTTTGC GGCTGTAGGA CCTGTTTTGC AAGGTTTAAA	19380
CAAGCCAGTT AATGATCTTT CTCGTGGATG TAATGCAGAT GATGTTTACA AGTTGACCTT	19440
CATCACAGCA GCTCAAGCAG TTCATCAATA GTGAAAATA TAAAGTGATA TACTATGCTA	19500
TACTGTAGTT ATGAACTAT GTACGAAAAG CACTGCCATT AATTCCTGAG AACTAAATTA	19560
CTGATTGGTG TCAAAAAGGA AAACCTCCAA GCGATGATAT CCTGTCTATA CACGACCTAT	19620
AGAAATCTGT AATATACATA TCCGTAAAAC GATAAATCC CTTTTTGATT TTAAATGAGT	19680
ATGAAAAGAG AATTTTTTGG CTCTTTGTCA ACTGTAGTGG GTTGAAGAAA AGCTAAGCTC	19740
GAGAAAGGAC AAATTCATC CTTTCTTTTT TGATATTCAG AGCGATAAAA ATCCGTTTTT	19800
TGAAGTTTTC AAAGTTCCGA AAACCAAAGG CATTCGCTT GATAAGTTTG ATGAGATTAT	19860
TGGTCGCTTC CAGTTGGCG TTAGAATAGT GTAGTTGAAG GGCGTTGATA ATCTTTTCTT	19920
TATCTTTGAG GAAGGTTTTA AAGACAGTCT GAAAAATAGG ATGAACCTGC TTAAGATTGT	19980
CCTCAATAAG TCCGAAAAAT TTCTCTGGTT CCTTATCTG GAAGTGAAAA AGCAAGAGTT	20040
GATAGAGCTG ATAGTGGTGT TTCAAGTCTT CCGAATAGCT CAAAAGCTTG TTTAAATCT	20100
CTTTATTGGT TAAGTGCATA CGAAAAATAG GACGATAAAA TCGCTTATCA CTCAGTTTAC	20160
GGCTATCCTG TTGAATGAGT TTCCAGTAGC GCTTGATAG	20199

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19702 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ACCCGATGTA TCAGCGGATA TTTACTCTAT TTTTCAAACG ATGTTATACC CACAATAAAA	60
GAAAAAAGAC CCTAAGGTCT CCTTGCTTT TATTATTAAA CGCGTTCAAC TTTACCTGAT	120
TTCAAAGCAC GAGCTGAAGC CCAAACCTTT TTAGGTTTAC CATCGATAAG AACAGTAACT	180
TTTTGAAGGT TTGGTTTTAC GGCACGTTTT GTTTGGTTCA TCGCGTGTGA ACGGTTGTTT	240
CCTGATACAG TCTTACGACC TGTAAGTAA CATACTTTAG CCATTGTGTT TTCCTCCTAT	300
TAGATCTAAT ATAGCGGATG TGCTAGCACC ACATACCGTA CTATGTTATC ACATTTTCTT	360
GTTTTTTGCA AGGGAATTGG AAGATTTTT ATTTGTGTCT TAAATCAGGT CTGCGTGAC	420
ATTTCTGCTC TCCACATGCC ATCGTTGATT AACAGAACAC CAGAATTAAA ATTATGTGTA	480
TAAAAATCAT CTCTAACTGC AGCTAAGGGT ATAGCCGTCA AGTCCAAATC CCACAGCTCA	540

195

TCTATCGATT TTCTTACAAC AATATCTGAA TCCAAATACA GTACACGAGA CTCGCTTACA	600
TACTTTGGAA TAAAATACCT AAAAAAGCCG CATATGAAAG TCCCTCAAAG GGGAGACGAT	660
AACCTTTCAG AATATTACTG TCAATCTAAA CATTCACAAT CTCACTATTC AAAGTCTCTA	720
GTCTTTTTC CATCAATTGG AACCATTCTC GCGGAAGGTC ATCATTAAAA ACATAAAACT	780
TAAGATTATA ATGATGAACA CAAAGAGATT TTATTGTTGT TTCAACTTTA TCCATATAAG	840
CATTATCTGC ACCTAAGACA ATCGCTTTT TCTCTTCTT CACTTTTAT CTCATTTCTT	900
TTTATTCCCA TCATATTATT CCCATCATAT GTTCCCATC ATATGTTTCT ACGTAACCAT	960
TATTTTCGCC TATTCGTTTCG TAAAACCATA CCAGTGGAGA TTTTAGATGA AGTCCCATTA	1020
CGGTTTACAA TTTTACATT ACGACACGGA GTTTTACAAA TCGATTTCAT TTGCCAAACG	1080
TAGTTAGTGA GGCAGTTAGC TAGTTCGCCA AATAGCGACT AGCGTCCAAC AATTTGGAAC	1140
TTTAGTTCCA ATTGTTGGTA CTGAGTCACA TCTTCTCCTC TAACTCTACG TCTGGATACT	1200
TGTCCGCAAA CCAGCGGAGG GCAAAGTCAT TTCAAAGAG AAAGACTGGT TGGTCAAAAC	1260
GGTCTTTGGC TAAGATATTG CGACTTGACG ACATCCGTTT ATCCAAGTCC TCAGGCTTGA	1320
TCCAACGAAC GGTCTTTTAA CCCATTGGGT TCATAACTAC TTCCGCATTG TACTCGCCTT	1380
CCATGCGGTG TTTAAAGACT TCAAACCTGGA GTTGACCTAC AGCGCCTAGC ATGTACTCAC	1440
CTGTTTGGTA ATTCTTATAA AGCTGAACGG CTCCTTCTTG CACCAATTGC TCAATCCCCT	1500
TGTGGAAGGA TTTTGTCTC ATAACATTCT TAGCAGAAAC TTTCATGAAA ATCTCAGGTG	1560
TAAAGTTGG CAGGGGTTC AATTCAAAC TGTTTTTTCC AACCGTCAAG GTATCCCCAA	1620
CCTGATAAGT ACCGGTATCG TAAACCCGA TAATATCACC TGCCACGGCA TTGGTCACAT	1680
TCTCACGACT CTCCGCCATA AACTGGGTAA CATTAGATAG TTTAGCCCC TTACCAGTAC	1740
GAGGGAGATT GACACTCATG CCGCGCTCAA ATTCGCCAGA TACGATACGG ACAAAGGCAA	1800
TACGGTCACG GTGACGAGGG TCCATGTTGG CTTGGATTTT AAAGACAAAG CCTGAGAAAT	1860
CCTTGTGATA AGGATCCACA ATTTACCGT CTGTTTCTT GTGACCATGT GGTCTGGAG	1920
CAAACCTGAG GAAGGTTTCA AGGAAGGTCT GCACACCAA GTTTGTGAGG GCTGAACCGA	1980
AAAAGACAGG CGTCAATTCT CCAGCCAGAA TAGCTTCCTC TGAAACTCA TTCCCGGCTT	2040
CATTTAAAAG CTCAATGTCA TCCTTGACTT GCTCGTAGAA AGGATTGCTA CCAAAGAGTT	2100
TGTCCCCGTC TTCTAGACTG GCAAAACGCT CATCCCCTT GTAAAGCTCT AAACGTTGGT	2160
TATAGAGGTC ATACAAGCCC TCAAAGGCTT TCCCATCCC GATAGGCCAG TTCATAGGGT	2220
AGCTAGCAAT GCCCAAGATT TCTTCCAATT CTGCAAGAG ATCCAAAGGC TCACGACCGT	2280

196

CACGGTCCAG CTTGTTCATA AAGGTAAAGA CTGGAATGCC ACGATGTTTC ACAACCTCAA	2340
ACAATTTCTT GGTTCGAGCC TCGATCCCCT TGGCAGAGTC CACGACCATG ACCGCAGCAT	2400
CCACCGCCAT CAAGGTACGA TAGGTATCTT CTGAGAAGTC CTCGTGCCCT GGCCTGTCTA	2460
AGATATTCAC GCGCTTGCCG TCGTAGTCAA ATTGCATAAC AGATGAAGTA ACAGAAATCC	2520
CACGTTGCTT CTCGATATCC ATCCAGTCAG ATTTAGCAAA AGTCCCTGTT TTCTTCCCTT	2580
TTACCGTACC AGCCTCACGA ATCTCACCCC CAAAGTAGAG TAACTGCTCA GTGATGGTTG	2640
TTTTCCCCGC GTCCGGGTGG GAGATAATGG CAAAGGTACG ACGTTTCTTA ATTTCTTCTT	2700
GAATATTCAT AAGTCTCTT TCTTTGATTC TCTATTTTC TTGTTTCAAT AGCTGAGAAT	2760
GATTTTACA TTGGATTTTA CCATTCCTTT CAACACTCCA TTATATCGGA TTTTAGCATT	2820
TTTTTCAATT TCTATTTCTT TTCACTTCCC CCTCCCTTAT TTATAGGAAA ATATGGTAAA	2880
ATAGAACAGA CTAATAATCA TCATTTACAG AAAGGATGCA AGATGAAAAT TACGCAAGAA	2940
GAGGTAACAC ACGTTGCCAA TCTTTCAAAA TTAAGATTCT CTGAAGAAGA AACTGCTGCC	3000
TTTGCGACCA CCTTGTCTAA GATTGTTGAC ATGGTTGAAT TGCTGGGCGA AGTTGACACA	3060
ACTGGTGTG CACCTACTAC GACTATGGCT GACCGCAAGA CTGTACTCCG CCCTGATGTG	3120
GCCGAAGAAG GAATAGACCG TGATCGCTTG TTTAAAAACG TACCTGAAAA AGACAACTAC	3180
TATATCAAGG TGCCAGCTAT CCTAGACAAT GGAGGAGATG CCTAATGACT TTTAACAATA	3240
AAACTATTGA AGAGTTGCAC AATCTCCTTG TCTCTAAGGA AATTTCTGCA ACAGAATTGA	3300
CCCAAGCAAC ACTTGAAAAT ATCAAGTCTC GTGAGGAAGC CCTCAATTCA TTTGTCACCA	3360
TCGCTGAGGA GCAAGCTCTT GTTCAAGCTA AAGCCATTGA TGAAGCTGGA ATTGATGCTG	3420
ACAATGTCCT TTCAGGAATT CCACTTGCTG TTAAGGATAA CATCTCTACA GACGGTATTC	3480
TCACAACTGC TGCCTCAAAA ATGCTCTACA ACTATGAGCC AATCTTTGAT GCGACAGCTG	3540
TTGCCAATGC AAAAACCAAG GGCATGATTG TCGTTGGAAA GACCAACATG GACGAATTTG	3600
CTATGGGTGG TTCAGGTGAA ACTTCACACT ACGGAGCAAC TAAAAACGCT TGGAACCACA	3660
GCAAGGTTCC TGGTGGGTCA TCAAGTGTT CTGCCGAGC TGTAGCCTCA GGACAAGTTC	3720
GCTTGCTACT TGGTTCTGAT ACTGGTGGTT CCATCCGCCA ACCTGCTGCC TTCAACGGAA	3780
TCGTTGGTCT CAAACCAACC TACGGAACAG TTTACGTTT CGGTCTCATT GCCTTTGGTA	3840
GCTCATTAGA CCAGATTGGA CCTTTTGCTC CTA CTGTTAA GGAAAATGCC CTCTTGCTCA	3900
ACGCTATTGC CAGCGAAGAT GCTAAAGACT CTACTTCTGC TCCTGTCCGC ATCGCCGACT	3960
TTACTTCAAA AATCGGCCAA GACATCAAGG GTATGAAAAT CGCTTTGCCT AAGGAATACC	4020
TAGGCGAAGG AATTGATCCA GAGGTTAAGG AAACAATCTT AAACGCGGCC AAACACTTTG	4080

197

AAAAATTGGG TGCTATCGTC GAAGAAGTCA GCCTTCCTCA CTCTAAATAC GGTGTTGCCG	4140
TTTATTACAT CATCGCTTCA TCAGAAGCTT CATCAAACCTT GCAACGCTTC GACGGTATCC	4200
GTTACGGCTA TCGCGCAGAA GATGCAACCA ACCTTGATGA AATCTATGTA AACAGCCGAA	4260
GCCAAGGTTT TGGTGAAGAG GTAAAACGTC GTATCATGCT GGGTACTTTC AGTCTTTCAT	4320
CAGGTTACTA TGATGCCTAC TACAAAAAGG CTGGTCAAGT CCGTACCCCTC ATCATTCAAG	4380
ATTTGAAAA AGTCTTCGCG GATTACGATT TGATTTTGGG TCCAACGCT CCAAGTGTTG	4440
CCTATGACTT GGATTCTCTC AACCATGACC CAGTTGCCAT GTACTTAGCC GACCTATTGA	4500
CCATACCTGT AAACCTGGCA GGACTGCCTG GAATTTGAT TCCTGCTGGA TTCTCTCAAG	4560
GTCTACCTGT CGGACTCCAA TTGATTGGTC CCAAGTACTC TGAGGAAACC ATTTACCAAG	4620
CTGCTGCTGC TTTTGAAGCA ACAACAGACT ACCACAAACA ACAACCCGTG ATTTTGGAG	4680
GTGACAACTA ATGAACCTTG AAACAGTCAT CGGACTTGAA GTCCACGTAG AGCTCAACAC	4740
CAATTCAAAA ATCTTCTCAC CTACTTCTGC CCACTTTGGA AATGACCAAA ATGCCAACAC	4800
TAACGTGATT GACTGGTCTT TCCCAGGAGT TCTACCAGTT CTCAATAAAG GGGTTGTTGA	4860
TGCCGGTATC AAGGCTGCTC TTGCCCTCAA CATGGACATC CACAAAAAGA TGCACCTTGA	4920
CCGCAAGAAC TACTTCTATC CTGATAACCC CAAAGCCTAC CAAATTTCTC AGTTTGATGA	4980
ACCAATCGGA TATAATGGCT GGATTGAAGT CAAACTAGAA GACGGTACGA CCAAGAAAAT	5040
CGGTATCGAA CGTGCCCAACC TAGAGGAAGA CGCTGGTAAA AACACCCATG GTACAGATGG	5100
CTACTCTTAT GTTGACCTCA ACCGCCAAGG GGTTCCTTG ATTGAGATTG TATCTGAGGC	5160
AGATATGCGT TCTCCTGAAG AAGCCTATGC TTATCTGACA GCCCTCAAGG AAGTTATCCA	5220
GTACGCTGGC ATTTCTGACG TTAAGATGGA GGAAGGTTG ATGCGTGTGG ATGCCAACAT	5280
CTCCCTTCGT CCTTATGGTC AAGAGAAATT CGGTACCAAG ACTGAATTGA AGAACCTCAA	5340
CTCCTTCTCA AACGTTGTA AAGGTCTTGA ATACGAAGTC CAACGCCAGG CTGAAATTCT	5400
TCGCTCAGGT GGTCAAATCC GCCAAGAAAC ACGCCGTAC GATGAAGCGA ATAAAGCAAC	5460
CATCCTCATG CGTGTCAAGG AAGGGGCTGC TGACTACCGC TACTTCCCAG AACCAGACCT	5520
ACCCCTCTTT GAAATTTCTG ACGAGTGGAT TGAGGAAATG CGGACTGAGT TGCCAGAGTT	5580
TCCAAAAGAA CGTCGTGCGC GTTATGTATC TGACCTTGGT TTATCAGACT ACGATGCTAG	5640
TCAGTTGACT GCTAATAAAG TCACTTCTGA CTTCCTTGAA AAAGCTGTTG CCCTAGGTGG	5700
TGATGCCAAA CAAGTCTCTA ACTGGCTCCA AGGGGAAGTC GCTCAGTTCT TGAATGCTGA	5760
AGGTAAACA CTGGAACAAA TCGAATTGAC ACCAGAAAAC TTGGTTGAAA TGATTGCCAT	5820

198

CATCGAAGAC GGTACTATTT CATCTAAGAT TGCCAAGAAA GTCTTTGTCC ATCTAGCTAA	5880
AAATGGCGGT GGC GCGCGTG AATACGTGA AAAAGCAGGT ATGGTTCAAA TTTCAGATCC	5940
AGCTATCTTG ATCCCAATCA TCCACCAAGT CTTTGCCGAT AACGAAGCTG CTGTTGCCGA	6000
CTTCAAGTCA GGCAAACGTA ACGCCGACAA GGCTTTACAG GATTCCTTAT GAAGGCAACC	6060
AAAGGCCAAG CCAACCCACA AGTTGCCCTT AAACACTTG CACAGGAATT GGCGAAGTTG	6120
AAAGAAAAC AGACAGAACA AAACCAGCCC TAAGGTGGT TTTTCTTCT CTACCAACTC	6180
CCAATAACTA TTTTGGCTTT ATTTCCAGAG TATTTTATGG TAAATGAAG AGTAATAATA	6240
TTTATTAAAG AGGTAAAAAC ATGATTGAAG CAAGTACCTT AAAAGCTGGT ATGACCTTTG	6300
AAACAGCTGA CGGCAATTG ATTCGCGTT TGGAAGCTAG TCACCACAAA CCAGGTAAAG	6360
GAAACACGAT CATGCGTATG AAATTGCGTG ATGTCCGTAC TGGTTCTACA TTGACACAA	6420
GCTACCGTCC AGAGGAAAAA TTTGAACAAG CTATTATCGA GACTGTCCCA GCTCAATACT	6480
TGTACAAAAT GGATGACACA GCATACTTCA TGAATACAGA AACTTATGAC CAATACGAAA	6540
TCCCTGTAGT CAATGTTGAA AACGAATTGC TTTACATCCT TGAAAACTCT GATGTGAAAA	6600
TCCAATTCTA CGGAAC TGAA GTGATCGGTG TCACCGTTCC TACTACTGTT GAGTTGACAG	6660
TTGCTGAAAC TCAACCATCT ATCAAAGGTG CTAAGTTAC AGGTTCTGGT AAACCAGCAA	6720
CGATGGAAAC TGGACTTGTC GTAAACGTT CAGACTTCAT CGAAGCAGGA CAAAACTCG	6780
TTATCAACAC TGCAGAAGGA ACTTACGTT CTCGTGCCTA ATCTCTAGAA AGAGGTCATT	6840
CTATGGGAAT TGAAGAACA CTTGGCGAAA TCGTTATCGC CCCACGTGTA CTGAAAAAA	6900
TCATTGCTAT CGCTACTGCA AAGGTAGAGG GTGTTCACTC TTTTCAAAC AGATCAGTGT	6960
CTGATACCCT TTCAAACTT TCACTCGGCC GTGGCATT TAATTAAAAAC GTGGACGAAG	7020
AACTCACAGC AGATATCTAT CTCTACCTTG AGTACGGAGT AAAAGTTCCT AAGGTAGCGG	7080
TTGCTATCCA GAAAGCTGTC AAAGATGCCG TCCGTAATAT GGCTGATGTA GAACTCGCTG	7140
CTATCAATAT TCACGTTGCA GGTATCGTCC CAGATAAAAC ACCAAAACCA GAATTGAAAG	7200
ATCTATTTGA CGAGGACTTC CTCAATGACT AGTCCACTAT TAGAATCTAG ACGCCAACTC	7260
CGTAAATGCG CTTTCAAGC TCTCATGAGC CTTGAGTTG GTACGGATGT CGAAACTGCT	7320
TGTCGTTTCG CCTATACTCA TGATCGTGAA GATACGGATG TACAACTTCC AGCCTTTTGT	7380
ATAGACCTCG TTTCTGGTGT TCAAGCTAAA AAGGAAGAAC TAGATAAGCA AATCACTCAG	7440
CATTTAAAAG CAGGTGGAC CATTGAACGC TTAACGCTCG TGGAGAGAAA CCTCCTTCGC	7500
TTGGGAGTCT TTGAAATCAC TTCATTTGAC ACTCCTCAGC TGGTTGCTGT TAATGAAGCT	7560
ATCGAGCTTG CAAAGGACTT CTCCGATCAA AAATCTGCCC GTTTTATCAA TGGACTGCTC	7620

AGCCAGTTTG TAACAGAAGA ACAATAAGGC TCTTTGTCAA CTGTAGTGGG TTGAAAAAAA	7680
GCTAAGCTCG AGAAAGGACA AATTTTCGTCC TTTCTTTTTT GATGTTCAAA GCGATAAAAA	7740
TCCGTTTTTT GAAGTTTTCA AAGTTTCGAA AACCAAAGGC ATTGCGCTTG ATAAGTTTGA	7800
TGAGATTATT GGTGCTTCC AGTTTGGCAT TAGAATAGTG TAGTTGAAGG GCGTTGACAA	7860
TCTTTTCTTT ATCTTTGAGG AAGGTTTTAA AGACAGTCTG AAAAATAGGA TGAGCCTGCT	7920
TAAGATTGTC CTCAATAAGT CCGAAAAATT TCTCTGGTTC CTTATTCTGG AAGTGAAACA	7980
GCAAGAGCTG ATAGAGCTGA TAGTGGTGTT TCAAGTCTTG TGAATGGCTC AAAAGCTTGT	8040
CTAAATCTC TTTATTGGTT AAGTGCATAC GAAAAGTAGG ACGATAAAAT CGCTTATCAC	8100
TCAGTCTACG GCTATCCTGT TGAATGAGTT TCCAGTAGCG CTTGATATCC TTGTATTAT	8160
GGGATTTTCG ATGAACTGA TTCATGATTT GGACACGCAC ACGACTCATG GCACGGCTAA	8220
GATGTTGTAC AATGTGAAAG CGATCAAGAA CGATTTTAGC ATTCGGGAGT GAAACAGTCT	8280
GGGAGACTGT TTCAGCCTGA GCCTAGGAAT TTGAAAGCGA AGCTGTTTAG CCAAGTCATA	8340
GTAAGGGCTA AACATATCCA TAGTAATAAT TTTGACGCGA CATCGGACAA CTCTATCGTA	8400
GCGAAGAAAG TGATTTTCGAA TGATAGCTTG TGTTCACCC TCAAGAACAG TGATGATATT	8460
GAGATTGTTA AAATCTGCG CAATGAAGCT CATCTTTCCC TTTGTAAAAG CATACTCATC	8520
CCAAGACATA ATCTCAGGAA GACAAGAAAA ATCATGTTTA AAGTGAAAAT CATTGAGCTT	8580
ACGAATAACA GTTGAAGTTG AGATGGAAAG CTGATGGGCA ATATCAGTCA TAGAAATCTT	8640
TTCAATCAAC TTTTGAGCAA TCTTTTGTTT GATGATACGA GGGATTGGT GATTTTCTT	8700
GACGATAGAA GTTTCAGCGA CCATCATTTT TGAACAGTGA TAGCACTTGA ATCGACGCTT	8760
TCTAAGGAGA ATTCTAGTAG GCATACCAGT CGTTTCAAGA TAAGGAATTT TAGAAGGTTT	8820
TTGAAAGTCA TATTTCTTCA ATTGGTTTCC GCACTCAGGG CAAGATGGGG CGTCGTAGTC	8880
CAGTTTGGCG ATGATTTCCT TGTGTGTATC CTTATTGATG ATGTCTAAAA TCTGGATATT	8940
AGGGTCTTTA ATGTCTAGTA ATTTTGTGAT AAAATGTAAT TGTTCATAT GAATCTTTCT	9000
AATGAGTTGT TTTGTCGCTT TTCATTATAG GTCATATGGG ACTTTTTTTC TACAATAAAA	9060
TAGGCTCCAT AATATCTATA GGGGATTTAC CCACTACAAA TATTATAGAG CCAACAATAA	9120
AAAGAAAAAG TGTTCGATAG ATATCAAACA CTTTTTCTT TGCCTCCAC TATCTAAAAA	9180
AATGATAATA GATATAATTG TAAACAAAAA TCCAGATAGG TTTTGCATGA TTGAGAAAGT	9240
TAAAAAACT ATGGCAGAGA ATCGTTAATC TCAGATTGTC GGTAGAACGA TAAACAAGGG	9300
CAAAAAAGAA ACCAATCAGA CTATAATATA ATAACTAAT TGGATCTCTG TGAGATAGTA	9360

200

TCAAATGGCT AATCCCAAAG ATGATAGCAG ATAGGATAAC ATCCAAATAG TACTTGGA	9420
AGGGAAAGAA GGTATTCATA AAATACCCTC TATCAAGAGT CTCCTCAAAA ACAGGACCGA	9480
TGATTACAGG CAGGACAAAA GATAAGATAG TCGATAAAAA GGTGGTGTGT CCATTTGAAA	9540
AAAGCACGGT AAAATACTCA TCATGAATAT TCCTATGATT AATCAAATGA GCATAGCGTG	9600
CCCCAAAATT ACCGAGAATC TGATAAACCA CATAAGTTGC AAATAAGTAG AAGACAAATG	9660
ACCAGTTCCA GCTCTTTTTC TCAAAGATAA AGAGCATCTT TTTCTTTTTC AACCTCCAAA	9720
TTAATAGAAG GAAACTTCCC ACTAATCCCA TTGTAAAT AAGAGAATAG ACATCAGCTC	9780
CTAACCCTAA AATGATCGTC ACATACAATC CAATTGTTTG TGGTAAATAG GTAGATAGTA	9840
AAATAATAAG CAAAATATT CCAAATTGTC TTAGTTTTTT TGTGTTTCTC ATCGTACTTT	9900
TTTGAAAGAT TACCCTGCTC GGAAGCCGTA CTCCAAGCA TCTATATAAG AATTAAGTGC	9960
CCCTTGCCTC ATATAGGGAG CAAATPCTCT ATAATATAAC CATCTACTAT ATCCATCTTC	10020
CCAAACAGCA AGACCACCTG AAGTTTGCTC CAAGTCTCA GTTGAAAGAA CTGTAAATGT	10080
ATTTGTACCT GTCATTGCAA GTACCTTCTT AAAATAGATT GTGTAGGCT CACATTTATA	10140
GTATATTCT TTTTTGTCT ATTTTATAGC CCATCTCCTC AACTGGCAAT TTTTCGACCT	10200
GAATTACATT TTTCCATAAA AAATGAGACC TTTCTAGTCT CATTTAGTCA TTCTTAGTAT	10260
TTTCTAAATC GTTGATAGCG TTCTTCCAGC AACTCTTCTA GCGGTTTTTG TGAAAGTCTA	10320
GCCAGCTCCG TTGGAGTTC TTTTTTGACA CTCTTAATCA GTTCTTTACT AGAAAGTCTT	10380
ATTTCAGAAA TCACCTTATC CACCACGTCC ATTTCTAACA GTTCATGCGA AGTGATTTTC	10440
ATCAGTTCTG CTGCTTCCAT AGCGCGAGTA CCGTCCTTCC ATAAATGGA AGCAAAGCCT	10500
TCTGGACTGA GAATGGCATA GATAGAATTT TCCAGCATCC AGACACGGTC CGCGACAGCT	10560
AGAGCCAGAG CCCCCTGA ACCACCTTCA CCGATAATAA TGGCGATAAT AGGAACTTTC	10620
AGGTCACTCA TTTCATGAG ATGCGAGCG ATAGCTTCCC CTTGACCACG TTCTTCCGCT	10680
CCGACACCAG GATAAGCACC TGCTGTATTG ATAAAGGTCA CAACTGGACG GCCAAATTTTC	10740
TCAGCCTGTT TCATCAACCG CAGTGCCTTT CCGTAGCCTT CTGGATGTGG TTGGCCAAAA	10800
TTCCGTTTGA GGTGTCTTG CAACTCTTG CCTTTTTGGA TACCAACCAC TGTTACAGCT	10860
TGGTCTCCAA GCCAACCAAT ACCACCAACA ACTGCACCAT CATCACGAAA AGAACGGTCA	10920
CCATGTAATT GGATAAATTC ATCAAAAATG CTGTGCGAA AGTCCAAGGT TGTCAAGCGA	10980
CTCTGCTCAC GCGCTTCTCT GACTATTTTT GCAATATTC TCTAGGACTC CCTCCATGCA	11040
ATCTGACTAG GCTAGCAATC GTATCTGGTA AGTCTCTTCT TTTGACAATA GCATCCACAA	11100
AGCCATGTTT TAATAGGAAT TCTGCCTTTT GGAAATCCTC AGGCAAGCTT TCACGAACCG	11160

201

TATTTTCAAT CACACGACGC CCAGCAAAAC CAACCAAGCT CTGTGGTTCA GCCAGAATGA	11220
TATCGCCTTC CATAGCGAAA GAAGCTGTCA CACCACCAGT CGTTGGATCT GTCAAAATGG	11280
TCAGGTAAAA GAGACCAGCA TTGAATGGC GTTTAACCGC CGCAGAGATC TTAGCCATCT	11340
GCATGAGACT CATGATTCCT TCCTGCATAC GGGCTCCACC AGAGGCTGTG AATAGGACAA	11400
CTGGCAATTT TTCGACAGTC GCATACTCAA ACAAACGAGT GATTTTTTCA CCTACAACCG	11460
TACCCATAGA AGCCATGATA AAGTTAGAAT CCATAATCCC AAGAGCCACA GTCTGACCTT	11520
TAATAAGAGC AGTTCCTGTC ACAACGGCTT CATGCAGACC TGTTTTTTCA CGCATAGATG	11580
CCAGTTTCTT TTGGTAACCA GGGAAATGCA AGGGATCCTT GCTTTCAATC CCTGTAAACA	11640
ATTCTTTGAA GGTTCCTATA TCAATCGTCA AAGCCAAGCG TTCTTGGCA GAAATACGAA	11700
AGGTATAGCT ACAGTGCAGA CAGATACGTT CACTTCCAG ATCCTTCTGA TAGATGGTAT	11760
GCTTACAGCC TGGACACTGG GAAAATAATT CATCTGGAAC CTCTGGCTTA GCTTGAGGTT	11820
TTTCCCTAAC CGAACGATTG GGATTGATTC GAATATACTT ATCTTTTTTA CTAAATAGAG	11880
CCATTGATTC CCCTTTTCGG TTAAACTCT TAAAGTCATT TTATTCTTTT TCTTGATATT	11940
TAGGTAAGAA GGTTCCTATC AAGAAGGAAG TATCATAATC CCCAGCAATG ACATTGCGAT	12000
CTGAAATGAG GTCAAGCTGG AAATCTGCAT TGGTCTGCAC TCCTTCAATT TCTAATTCAT	12060
AGAGGGCAGC TTGCATTTTC ATCAAGGCGT CAAAACGATT TTCGCCGTGT ACTATGATTT	12120
TGGCAATCAT ACTATCATAA TAAGGCGGAA TGGTATAACC TGGATAAACT GCTGAATCCA	12180
CGCGCAAGCC AACTCCACCA CTTGGCAGAT AGAGATTAGT AATCTTACCT GGACTTGGAG	12240
CAAAGTTAAA GGCTGGGTTT TCTGCATTGA TACGACACTC GATGGCATGA CCGCGTAGGA	12300
CAATATCTTC TTGCTTAACA GACAAAGGCT GACCTGCCGC AATGCAAATC TGTTCCTTAA	12360
CGATATCAAC ACCTGAAACA AACTCTGTTA CTGGATGTTT TACCTGAACA CGAGTATTCA	12420
TCTCCATGAA ATAGAAATTG CTACTTGCTT CATCAAGAAG AAATTCAATG GTTCCTGCAT	12480
TCTCATAGCC AACAACTCT GCCGCTCGAA CAGCAGCAGC ACCTATTTCA TGACGCAGCG	12540
TTTTTCCGAT TGCAATCGAG GGACTTTCTT CCAAACCTT TTGGTTATTC CTTTGAAGAG	12600
AACAATCCCG TTCACCCAAG TGAATCACAT GTCCATGCTC ATCACCTAGG ATTTGAACCT	12660
CAATGTGCCG AGCTGGATAG ATAACCCGTT CTATGTACAT GGCACCATG CCATAATTGG	12720
CCTTGGCCTC ACTAGAGGCA GTTTCAAAGG CAGAAACGAG GTCATCTGGT TTTTCAACCT	12780
TACGAATCCC TTTACCACCT CCACCTGCTG AAGCCTTGAG CATAACAGGA TAGCCAATTT	12840
TTTCAGCAAC AATCAAAGCT TCTTCAGAGT TATGCACTTC TCCATCTGAA CCTGGTATAA	12900

202

CAGGCACACC TGCTTTAATC ATCTGAGCAC GCGCATTTGAT CTTATCCCCC ATCATATCCA	12960
TAACATGACC AGATGGACCG ATAAACTTGA TACCTACTTC TTCACACATG GTCGCAAATT	13020
TGGAATTTTC ACTGAGAAAT CCAAAACCAG GGTGAATAGC TTCTGCCTCA GTCAAGACTG	13080
CAGCTGATAG AACTGCATTA ATATTGAGAT AAGACTCTGT TGCCTTGCCA GGACCAATAC	13140
AAACTGCTTC ATCTGCCAAA AGCGTATGAA GAGCTTCCTT ATCAGCAGTT GAATAAACCG	13200
CTACCGTCGC AATCCCCAAT TCACGTGCCG CACGGATAAT ACGAACCGCA ATTTCAACCAC	13260
GATTGGCAAT TAAAATTTTT CGAAACATGG AGAACCTCCT TAGTTCCCAA TTGCAAAAGT	13320
AAGGGTACCA CTGGCTGCAA GCTTGCCATC CACTTCAGCC TTTGCTTCAA CCACAGCTAT	13380
GGTGCCACGA CGTTTACAA AAGTCGCTGT CATAACCAAT TGGTCGCTG GTACAACCTG	13440
CTTCTTGAAC TTAACCTTGT CCATACCAGC GTAAAAGACC AGTTTTCTTT TATTTTCAGG	13500
TTTGTATAAC TCCAACACAC CGGCAGTTTG CGCCAAGGCT TCCATAATCA CAACACCTGG	13560
CATAACTGGG TATTGAGGAA AGTGGCCGTT AAAGAAAGGC TCGTTGATGG TCACATTTTT	13620
GATAGCAACA ATGGTATCCT CGCTCACTTC CAAGACACGG TCCACTAGAA GCATAGGATA	13680
ACGGTGGGGA AGAGCTTCTT TGATTCTTG AATATCGATC ATTTGATACG TACCAATCCT	13740
TTACCAAACCT CAACCATTTT TTCGTTAGAG ACGAGAATTT CCGTTACCAC ACCATCCTTA	13800
GGAGCTGGGA TTTTCATTCAT GACTTTCATG GCTTCGATAA TTACCAATGT TTGACCTTTT	13860
TTGACACTAT CACCAACTGT AACGAAGGCA GGTTTATCTG GTCCAGCAGC CAAGTAAACC	13920
ACTCCAACAA GTGGACTCTC TACAAGATTT CCCTCAGTAG CCACACTTGC TTCAGCTGGA	13980
GCTGGAACCT CTTCTGCTAC AGTCTCTGCT GGAGCAGATG TAGGAGCTAC TGGACTCGGT	14040
GTGCTAGAA CGGGTGCTGG AGCGACTTGA GTTGCAACTT CAGGCACAGG TCTTGCTTCA	14100
TTCTTGCTAA ACTGCAACTC ATCCGTCCCA TTTTATAAG AAAATTCTCT CAAACTTGAC	14160
TGGTCAAATT GAGTCATCAA GTCTTTAATA TCGTTTAAAT TCATACTTAT CTATTCTCCC	14220
AACGTTTGAA AGCAAGAACT GCATTGTGGC CTCCAAAACC AAAAGTATTT GAAATAGCGT	14280
ATGGAATTTT TTTCTCCAAG CCTTGTCAT AAACGACATT AGCTTCGATA TAATCTGATA	14340
CTTCACTTGT CCCAGCTGTC ATTGGTACAA AGTTATGACG CATAGCTTCG ATGGTGACGA	14400
TAGCTTCTAC TGCACCCGCA GCCCCAGCA AATGTCCTGT AAAAGACTTG GTTGATGATA	14460
CAGGTACTTC CTTACCAAGA ACAGCTACGA TAGCACCCTT TTCTCCTTTT TCATTGGCAG	14520
GAGTTGACGT TCCGTGAGCA TTGACATAGG CTAATTGCTC TGGAGAAATC TCAGCTTCTT	14580
CCAAGGCTAG TTTGATGGCC TTGATAGCTC CCTGACCTTC TGGATGTGGA GAAGTCATGT	14640
GGTAGGCATC ACAAGTATTT CCGTAACCAA CCACTTCAGC CAGGATAGTA GCTCCACGTT	14700

TTTCAGCGTG TTCAAGACTT TCTAGAACCA ACATCCCTGA ACCTTCACCC ATAACAAACC	14760
CATTGCGATC CTTATCAAAT GGGATCGAAG CACGAGTTGG ATCCTCTGTA GTAGAGAGAG	14820
CTGTTAAGGC TTGGAAACCA GCGATGGCAA AAGGTGTGAT AGAAGCTTCT GTTCCTCCCA	14880
CCAACATCAC ATCTTGGAAC CCAAACCTAA TGGAGCGGAA GGCATCCCCA ATCGCATCAT	14940
TTGATGAAGA GCAGGCAGTA TTGATAGATT TACAAACACC GTTTGCACCA AAACGCATGG	15000
CTACATTCCC AGAAGCCATA TTTGGTAAAG CTTTGGGAAG AGTCATTGGT TTGACACGTT	15060
TGGGTCTTTT TTCATGAAGG CGAAGTACCT GATCTTCAAT TTCCTTGATT CCACCAATAC	15120
CAGATGCAAC GATAACACCA AAACGATCCC TATTAAGAGC CTCTACATCA AGATTGGCAT	15180
GATTTACAGC CTCTTGGGCT GCATACAAGG CATATAAAGA ATAGTTATCA AAACGGTTGG	15240
TATCTTTTTT TACAAAGTAT TTATCGAACG GAAAATCTTG GATTTCTGCC GCATTATGCA	15300
CATCAAAGTC ACTATGATCA AATTTGTAA TGCCACCAAT GCCGATTTTC CCAGTTGCTA	15360
AACTATTCCA AAATCTTCT GGTGTATTTC CGATTGGAGA TGTTACTCCA TAACCTGTTA	15420
CCACTACTCG ATTTAGTTTC ATTCTTTTCA CCTCTAGCTT TCGCTACATA CTTAAGCCAC	15480
CATCAATGGC AACCATTGT CCAGTTAGAT AATCTGGCC TGCTAAAAAT ACTGTCAAAT	15540
CTGCAACCTG CTCTGCCTGC CCAAATCTT TCATCGGAAT CTGAGCTAGT GTAGCTTCCT	15600
TAATCTTATC TGACAGGATA GCGGTCATAT CAGACTCAAT CATTCCTGGA GCAATCACAT	15660
TGACTCGTAT ATTCGGACTA GCGACCTCGC GTGCCACAGA CTTGGTAAAG CCAATCAAGC	15720
CAGCCTTAGA AGCAGCATAA TTAGCTTGAC CAATATTCCC CATCAAACCA ACAACACTAG	15780
ACATATTAAT GATAGCACCT TCTCTGGCTT TCATCATCGG TTTCAGACT GATGTGTGCA	15840
TATTAAAGGC ACCAGTCAGA TTGACCTTGA GCACTTTTTC AAAATCTGCT TCTGTCATCT	15900
TGAGCATAAG AGTATCTTGG GTAATCCCTG CATGTTGAC CAAAACATCT ACTGAACCCA	15960
GTTCGCAAT AGCTTGATCA ATCATACGCT TAGCGTCTGC AAAATCTGAT ACATCTCCTG	16020
AAATGGGAAC CACCTTGATA CCATAGTTTG AAAACTCAGC GAGCAATCT TCTGAGATTG	16080
CCCCACGACT GTTTAAGACA ATGTTGGCTC CTGCTTGAGC AAACCTGTGG GCGATGGCAA	16140
GACCAATTCC ACGACTCGAA CCTGTAATAA AGATATTTTT ATGTTCTAGT TTCATTTTTT	16200
TCCTTTCAA ACTTCTACTT ATTTTAGTCT ATTTTCTAA AAGTGCTACT AAACCTGCTT	16260
GATCTTCCAC ATGAGCTAAG TGAGCAGTTT GATCAATTTT TTAAACAAA CCTGACAAGA	16320
CTTTCCCCGG TCCAATCTCG ATAAAGTTGC TTATGCCTGC TTCTTGCAATG ACCCAATAC	16380
TTTCATAGAA ACGAACGGGT TCCTTGACCT GACGCGTCAA GAGCTGAGCA ATGTCCTCTT	16440

204

TTTGCATCAC AGCAGCTTCT GTATTGCCGA CTAGGGGACA AGTAAATCT GAAAACTTA	16500
CCTGAGCTAG AGTTTCAGCT AGTTTCTGGC TAGCAGGTTT AAGGAGAGCG GTGTGAAAGG	16560
GACCTGACAC CTTAAGAGGA ATCAAGCGTT TGGCACCTGC TTCTTGCAA AGTTCAACCG	16620
CTCGATCAAC TGCAACCACT TCTCCAGCAA TGACGATTG TGCAGGTGTG TTATAGTTGG	16680
CTGGAGTAAC CACTCCAAGT TCAGAAGCTT TTTGACAGGC TTCTTCAATG ACCTCTACTG	16740
GCGTATTGAG AACTGCTACC ATCTTGCCAG AGTCAGCAGG AGCCGCTTCT TCCATATAGG	16800
CTCCACGCTT AGCTACCAAG GCAACCGCAT CTTCAAAATC CAAGGCGCCA CTGCCCACCA	16860
AGGCAGAGTA TTCTCCAAGA GACAAACCAG CAACCATATC AGGCTGATAG CCCTTTTCTT	16920
GCAATAAACG GTAGATAGCA ACCGAAGTCG CTAGAATGGC TGGTTGCGTA TAGCGGGTCT	16980
GATTGAGTTT GTCTTCTTCC GTATCGATGA GATAACGCAA ATCATAACCG AGCACCTGGC	17040
TCGCTCGATC AATCGTTTCT TTAACAATCG GATACTGATC ATAGAAATCC CGTCCCATCC	17100
CTAGATACTG GGCACCTTGA CCAGCAAATA AAAAGGCTGT TTTAGTCATT TCTTACAACT	17160
CCTGTCCAGC GAGAGGCTTC TTCTTGAATT TTCTTAGCGG CTCCGTAATA CAAATCTTTT	17220
AGGATTCTCT CAGCTGTTTC TTCTTTAGAA ACAAGCCCTG CGATTTGACC TGCCATAACA	17280
GAGCCACCAT CCACATCACC GTGAACAACT GCTTTGGCTA GAGCACCTGC TCCCATTTGT	17340
TCAAAGATTT CTAAATCAGG ATCTTCTTGC TTAAAGGCAT CTTTTTCAGC CAGTTCAAAA	17400
TCTCTAGTCA ACTGATTTT AATAGCACGA ACAGCATGAC CAAAGTGCTG AGCTGAAATC	17460
GTAGTATCAA TATCCCTTGC TTTTAAATTT TTCTCCTTGT AGTTTGATG GGCATTGAC	17520
TCTTTTGCAA CTACAAACCG TGTCCCCACC TGTACAGCCT CTGCACCTAG CATAAGCCA	17580
GCCGCAGCAC CTTCAACATC CGCAATTCCT CCTGCAGCAA TAACAGGAAT AGATATAGCT	17640
GTGGCTACCT GTCGCACCAA GGTGATGTT GTTAATTTAC CGATATGCCC CCCAGCTTCC	17700
ATTCCCTCTG CAATAACAGC GTCTGCACCG ATTTTTCCTA TGCCTTTAGC TAAAGCGACA	17760
CTAGGAACAA CAGGAATAAC GATTATCCCA GCTTCATGGA AACGTTCCAT ATACTTGCTT	17820
GGATTTCTG CTCCTGTTGT GACAACTTTA ACACCTTCTT CAATAACGAG ATCCACGATG	17880
TCTTCCACAA AGGGAGATAA GAGCATGATG TTGACCCCAA AGGGTTTATC AGTCAATGAT	17940
TTGATTTTAT CAATATTGGC CTTGACAACT TCTTTCGGGG CATTTCCCC ACCGATAATT	18000
CCTAATCCTC CAGCCTTGGA AACAGCCCCT GCCAAATCAC CATCAGCAAC CCAGGCCATC	18060
CCTCCTTGGA AAATAGGATA ATCAATCTTC AATAATTCTG TAATACGCGT TTTTATAGTG	18120
CCTCCAACCT TCCTTGCTTA CGTAATAGTT CGATTTTACC ATAATTGAC AGTCAAACTA	18180
TTACCTAAAC AAGAGGGAGT GGTTTCTCC CTACTCCTTC TACTAATATT CTGCTTATTT	18240

205

TGCTTGCTCT TCAACGTAAG CAACCAAGTC ACCAACTGTT TTCAAGTCAT TTTCTGCTTC	18300
GATTTGGATA TCAAAAGCAT CTTGATTTC TGAGATTACT TGGAACAAGT CCAATGAATC	18360
TGCGTCCAAA TCATCAAAAG TTGATTCAAG TGTTACTTCT GATGCGTCTT TTCCAAGTTC	18420
TTCAACGATA ATTTCTTGTA CTTTTTCAAA TACTGCCATG ATAGGACTCC TTTAAAATAA	18480
ATAGTTTTTT TATAACAATG TGTTACCAC ATGATTACCT AAATTGTAAG AATGAGCGTG	18540
CCCCAGGTCA AGCCTCCACC GAAGCCTGAT AGAAGAACAG TCTGGCTACC ATCTAAAGGG	18600
ATGAGACCTT GTTCTACACA CTCTGAAAGT AAAATCGGGA TACTGGCTGC ACTGGTATTG	18660
CCATATTCCA TCATATTGGC TGGAAGTTTG GTCGGTCAA CACCAATTTT TCTAGCCATC	18720
TTATCCAAAA TACGGTCATT GGCTTGATGA AGTAGCAGAT AATCCAAGTC TGTCACCTCT	18780
ATAGGAGATT CATCAATAGT CTGCTTGATA GACTTGGCTA CATCTCGAAT GGCAAAATCA	18840
AAGACTGTGC GTCCATCCAT CTTCAAAAAC GAATCTGCAC TTTCTTGATC TGAAAATGGA	18900
GAATGTAAAC CTGAATGCCC ATAAGTTAAA CACTCGCTGC GACTTCCATC GCTATTGAGA	18960
CTCTCAGCTA AGAAATGCTC TTGCTCGCTA GCTTCTAACA AGACACCACC AGCACCATCT	19020
CCAAACAACA CAGCTGTTGA TCGATCCGAC CAATCGACTG CCTTAGAGAG GGTTCCTACTA	19080
CCAATCACC AAGCCTTTTG AAAGCGACCA GAAGCGATAA ACTTTTCAGC AGTTGAAAGA	19140
GCAAAATACAA ATCCACTGCA AGCCGCGGTT AAGTCAAAAG CAAAGGCTTT ATTAGCACCA	19200
ATATTAGCTT GAACACGAGC AGCTGTAGAG GGCATCATCG AATCTGGAGT AATGGTAGCT	19260
AGGATGATAA AATCCAGTTC TTCTCCTGTT ATTCCAGCTT TTGCCATCAG TTTCTTAGCA	19320
ACCTCTGTAG CCAAATCACT GGTAGATTCT GTTCTTGAAA TATGCCTTTG TCGTATTCCC	19380
GTTCGACTTG AAATCCACTC ATCATTTGTA TCCATAATCT GAGCCAAGTC GTGATTTGTA	19440
ACCACTTGCT CTGGCACATA ATGAGCAACC TGACTTATTT TTGCAAAAGC CATTATTTCA	19500
AATCCTCCAA AAATTGGTAA AGATTAGTCA AACCTTTACC CATGACAGCA ATTTCTTCCT	19560
CGCTCATGCC ATCAATAATT TTTTCTACCA TGGCCTTGTG GAAGCGTTTA TGCAGTCTAT	19620
GAATCAAGCG ACCCTTCTTT GTCAAATGCA GATGCACCAC ACGACGATCC TGTCTGACC	19680
GAACTCGCTC AATGTAGCCC GG	19702

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6211 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

206

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GAAAAATTCC TCTCTTCTCT TGAAAAATTT TGA AAAAATG GTATGATAGT AACAAAGTTAT	60
TTTAAAGAGG AAAGAAAGGG GAATAATGGA GAAAATCAGT TTAGAATCTC CTAAGACGGG	120
GTCGGACCTA GTTTTGAAA CACTTCGTGA TTTAGGAGTT GATACCATCT TTGGTTATCC	180
TGGTGGTGCG GTTTTGCTT TTTATGATGC GATATATAAT TTAAAGGCA TTCGCCACAT	240
TCTAGGGCGC CATGAGCAAG GTTGTTCGA TGAAGCTGAA GGTATGCCA AATCAACTGG	300
AAAGTTGGGT GTTGCCGTCG TCACTAGTGG ACCAGGAGCA ACAAATGCCA TTACAGGGAT	360
TGCGGATGCC ATGAGCGATA GCGTCCCTT TTTGGTCTTT ACAGGTCAGG TGGCGCGAGC	420
AGGGATTGGG AAGGATGCCT TTCAGGAGGC AGACATCGTG GGAATTACCA TGCCAATCAC	480
TAAGTACAAT TACCAAGTTC GTGAGACAGC TGATATCCG CGTATCATTA CGGAAGCTGT	540
CCATATCGCA ACTACAGGCC GTCCAGGCC AGTTGTAATT GACCTACCA AAGACATATC	600
TGCTTTAGAA ACAGACTTCA TTTATTCACC AGAAGTGAAT TTACCAAGTT ATCAGCCGAC	660
TCTTGAGCCG AATGATATGC AAATCAAGAA AATCTGAAG CAATTGTCCA AGGCTAAAAA	720
GCCAGTCTTG TTAGCTGGTG GTGGAATTAG TTATGCTGAG GCTGCTACGG AACTAAATGA	780
ATTTGCAGAA CGCTATCAAA TTCCAGTGGT AACCACTCTT TTGGGACAAG GAACGATGTC	840
AACGAGTCAC CCACTCTTTC TTGGAATGGG AGGCATGCAC GGGTCATTTC CAGCAAATAT	900
TGCTATGACG GAAGCGGACT TTATGATTAG TATGGTTCT CGTTTCGATG ACCGTTTGAC	960
GGGGAATCCT AAGACTTTCG CTAAGAATGC TAAGGTTGCC CACATTGATA TTGACCCAGC	1020
TGAGATTGGC AAGATTATCA GTGCAGACAT TCCTGTAGTT GGAGATGCTA AGAAGGCCTT	1080
GCAAATGTTG CTAGCAGAAC CAACAGTTCA CAACAACACT GAAAAGTGGA TTGAGAAAGT	1140
CACTAAAGAC AAGAATCGTG TTCGTTCTTA TGATAAGAAA GAGCGTGTGG TTCAACCGCA	1200
AGCAGTTATT GAACGAATTG GTGAATTGAC GAATGGAGAT GCCATTGTGG TAACAGACCT	1260
TGGTCAACAC CAAATGTGGA CAGCTCAGTA TTATCCCTAC CAAATGAAC GTCAGTTAGT	1320
GACTTCAGGT GGTTCGGGAA CAATGGGCTT TGGAAATCCA GCAGCAATCG GTGCTAAAAT	1380
TGCTAACCCA GATAAGGAAG TAGTCTTGT TGTGGGGAT GGTGGTTTCC AAATGACCAA	1440
CCAGGAGTTG GCTATTTTGA ATATTTACAA GGTGCCAATC AAGGTGGTTA TGCTGAACAA	1500
TCATTCACTT GGAATGGTTC GCCAGTGGCA GGAATCCTTC TATGAAGGCA GAACATCAGA	1560
GTCGGTCTTT GATACCTTC CTGATTTCCA ATTGATGGCG CAGGCTTATG GTATTA AAAA	1620
CTATAAGTTT GACAATCCTG AGACCTTGGC TCAAGACCTT GAAGTCATCA CTGAGGATGT	1680

TCCTATGCTA ATTGAGGTAG ATATTTCTCG TAAGGAACAG GTGTTACCAA TGGTACCGGC	1740
TGGTAAGAGT AATCATGAGA TGTTGGGGGT GCAGTTCCAT GCGTAGAATG TTAACAGCAA	1800
AACTACAAAA TCGTTCAGGA GTCCTCAATC GCTTTACAGG TGTCCTATCT CGTCGTCAGG	1860
TTAATATTGA AAGCATCTCT GTTGGAGCAA CAGAAGATCC GAATGTATCG CGTATCACTA	1920
TTATTATTGA TGTGCTTCT CATGATGAAG TGGAGCAAAT CATCAAACAG CTCAATCGTC	1980
AGATTGATGT GATTGCGATT CGAGATATTA CAGACAAGCC TCATTGAG CGCGAGGTGA	2040
TTTTGGTTAA GATGTCAGCG CCAGCTGAGA AGAGAGCTGA GATTTTAGCG ATTATTCAAC	2100
CTTTCCGTGC AACAGTAGTA GACGTAGCG CAAGCTCGAT TACCATTGAG ATGACGGGAA	2160
ATGCAGAAAA GAGCGAAGCC CTATTGCGAG TCATTGCCCC ATACGGTATT CGCAATATTG	2220
CTCGAACGGG TGCAACTGGA TTACCCGCG ATTAATAATC CAACTTAAAT TTATTAAACC	2280
AGCCTAAAA GCAATAAATA ATAGAAAAGA GAGAAAAGCT ATGACAGTTC AAATGGAATA	2340
TGAAAAAGAT GTTAAAGTAG CAGCACTTGA CGGTAAAAAA ATCGCCGTTA TCGGTTATGG	2400
TTCAACAAGG CATGCGCATG CTCAAACTT GCGTGATTCA GGTGCTGACG TTATTATCGG	2460
TGTACGTCCA GGTAAATCTT TTGATAAAGC AAAAGAAGAT GGATTGATA CTTACACAGT	2520
AGCAGAAGCT ACTAAGTTGG CTGATGTTAT CATGATCTTG GCGCCAGACG AAATCAACA	2580
AGAATTGTAC GAAGCAGAAA TCGCTCCAAA CTTGGAAGCT GGAAACGCAG TTGGATTGTC	2640
CCATGGTTTC AACATCCACT TTGAATTTAT CAAAGTTCCT GCGGATGTAG ATGTCTTCAT	2700
GTGTGCTCCT AAAGGACCAG GACACTTGGT ACGTCGACT TACGAAGAAG GATTGCGTGT	2760
TCCAGCTCTT TATGCAGTAT ACCAAGATGC AACAGGAAAT GCTAAAAACA TTGCTATGGA	2820
CTGGTGTAAG GGTGTTGGAG CGGCTCGTGT AGGTCTTCTT GAAACAACCT ACAAGAAGA	2880
AACTGAAGAA GATTGTTTG GTGAACAAGC TGTACTTTGT GGTGGTTTGA CTGCCCTTAT	2940
CGAAGCAGGT TTCGAAGTCT TGACAGAAGC AGGTTACGCT CCAGAATTGG CTTACTTTGA	3000
AGTTCTTCAC GAAATGAAAT TGATCGTTGA CTTGATCTAC GAAGGTGGAT TCAAGAAAA	3060
GCGTCAATCT ATPTCAAACA CTGCTGAATA CGGTGACTAT GTATCAGGTC CACGTGTAAT	3120
CACTGAACAA GTTAAAGAAA ATATGAAGGC TGTCTTGGA GACATCCAAA ATGGTAAATT	3180
TGCAAAATGAC TTTGTAAATG ACTATAAAGC TGGACGTCCA AAATTGACTG CTTACCGTGA	3240
ACAAGCAGCT AACCTTGAAA TTGAAAAAGT TGGTCAGAA TTGCGTAAAG CAATGCCATT	3300
CGTTGGTAAA AACGACGATG ATGCATTCAA AATCTATAAC TAATTAGAAA TATATAGCGC	3360
TGGAGATGAT TTTATGAAAA AGATTATGAG AAAAATTGCA TCGTTATTAT TGGTTCTAGT	3420

208

TGTATAATGT AATTACACCG TCGGTAATAG TGCTAGCAGA CCAAAATAAA GCAGATTGGT	3480
CGTATGATGA AAATGCTGTA ATTAACATTT ATGATGATGC TAATTTTGAA GATGGTAGGT	3540
TGCATATGAA CTTTGAACAA TTCTTCAAAT TGGCACAAAT AGCTAGAGAA GAAGGCTTGT	3600
AAATTCATTC TCCGTTTGAG AGAGCTGGTG CCACTAAATC TGCTCGTTAT ATAGCGAAAT	3660
GGATTTTGAG AAATAAAAAA CATTAACAAA TATAGTTGGT AAATCATTAG GACCTAAATC	3720
AGCTGTTAGA TTCGGAGAAG CTTTATCCTA TATTGAAGGT CCTCTCGCA GAATAAATGA	3780
GACGATAGAT GGCGGTTTAT ATCAAATAGA GCAAAATATT GCATCTGGAT TGAAGAATC	3840
GGGTTTAAAT GACTGGACTG CGAAAACTTT AGCTTCAGCT ATTCGTGGGA TATTAGATGT	3900
ACTTATTTAG GGGTTGAAAT CATATGAATA TTACCAATTT GTTTTCTATC AAGACAGGAT	3960
GTGATGAAAC TGATAGGCAA CTGCAAAAAC TATTTTTC A GTTGGATTTA CAATTGGGAG	4020
AATTGACAGA TCAACTAAGA AAATTAGATT CTAATTTTGT TCCTCGTAGT CAATTTGTAG	4080
ACACGTTGGA TTTGAATGAT GTAGAATATA AAGAAATTTT AAATATTTT ATCTTCCATC	4140
GTAATGATAG TGAAGAAAGT TTGGTAGAAT GGTATATGA TTGGATTTC ACAAATCGTT	4200
ATGAACCTCC TAAAGAGTTT TCGATTCGTA TGGCTCATAA ATACCATGAA AGTGTACTG	4260
AAGTTTTCGG AGATGAATAA CTAAAAACA GTCATTAGTG ACTGTTTTT ATAGAAAAAG	4320
AGGTTTATA TGTTAAGTTC AAAAGATATA ATCAAGGCTC ACAAGGTCTT GAACGGTGTG	4380
GTGTGAATA CTCCACTGGA TTACGATCAT TATTTATCGG AGAAGTATGG TGCTAAGATT	4440
TATTTGAAAA AAGAAAATGC CCAGCGTGT CGCTCCTTTA AAATTCGTGG TGCCTATTAT	4500
GCCATTTCCC AGCTCAGCAA GGAAGAACGT GAACGTGGGG TAGTCTGCGC TTCTGCGGGA	4560
AATCATGCGC AGGGAGTAGC CTATACTTGT AATGAAATGA AAATTCCTGC TACTATCTTT	4620
ATGCCCATTA CTACGCCACA ACAAAGATT GGTCAAGTTC GCTTTTTTGG TGGGGATTTT	4680
GTAACATTA AACTAGTTGG AGATACCTTT GATGCCTCAG CCAAAGCAGC TCAAGAATTT	4740
ACAGTCTCTG AAAATCGTAC CTTTATGAT CCTTTTGATG ATGCTCATGT TCAAGCAGGT	4800
CAAGGAACAG TTGCTTATGA GATTTTAGAA GAAGCTCGAA AAGAATCGAT TGATTTTGAT	4860
GCTGTCTTGG TTCCTGTTGG TGGTGGCGGT CTCATTGCCG GGGTTTCTAC CTATATCAAG	4920
GAAACAAGTC CAGAGATTGA GGTATCGGA GTAGAGGCGA ATGGAGCGCG TTCCATGAAA	4980
GCTGCCTTTG AGGCTGGAGG TCCAGTAAAA CTCAAGGAAA TTGATAAATT TGCTGATGGG	5040
ATTGCTGTGC AAAAGGTAGG TCAGTTGACC TATGAAGCAA CTCGTCAACA TATTAAAACT	5100
TTGGTAGGTG TCGATAGGG ATTGATTTCT GAAACCTTGA TTGACCTTTA CTCTAAGCAA	5160
GGGATAGTCG CAGAACCTGC TGGAGCGGCT AGTATCGCCT CTTTAGAGGT TTTAGCTGAA	5220

209

TATATTAAGG GGAAAACCAT TTGTTGTATC ATTTCTGGAG GAAATAATGA TATCAACCGT	5280
ATGCCAGAAA TGGAAGAGCG TGCCTTGATT TATGATGGTA TCAAACATTA CTTTGTGGTC	5340
AATTTCCCAC AACGTCCAGG AGCTTTGCGT GAGTTTGTA ATGATATCCT GGGGCCAAAT	5400
GATGATATCA CACGTTTGA GTATATCAAA CGAGCTAGCA AGGGAACAGG CCCAGTATTA	5460
ATTGGGATCG CTTTAGCAGA TAAGCATGAT TATGCAGGTT TGATTCTAG AATGGAAGGT	5520
TTTGATCCAG CTTATATTAA CTTAAATGGT AATGAAACGC TTTATAATAT GCTTGTCTGA	5580
GGACTAATAA AAAAATATCA TACCTTCATT TTGATTTTCT ATCTATTGAC AAGCATAGTC	5640
ACACTGTCTT TAATACTCTT CGAAAATCTC TTCAAACCAC GTTAGCTCTA TCTGCAACCT	5700
CAAAACAGTG TTTTGAGCAA CTTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTCATTG	5760
AGTATAAGGT ATGATTGAT TTCTTTTGT TGACAAATAT ACTATATTAA AAAGATATAT	5820
AAGTAATTAA CTGAGCTTAT CTGTCTGTC ATCTCTATTA AGGATGCTTT AGATAATCGG	5880
GTGTCTGCTT CTAGGCTAGC ACCTCAATAT CCAAAGGAGT GATGAATTTG AAGGACATAA	5940
GGAATACCTA TCTCTCAGAT GATTTATTGA GGAAGAAAGA TAGGAGTTT TGAGCTAGTG	6000
AAGGCTTGA TTTCTAAAG TTAGAACTAT CATCTTCAGT TCTTAAATCG AAGAAATAAG	6060
CTATCTTACG GAAATAGAGA AGCATTTTT AAGAACTGA ATAATTTTCG ACCTTAAGAG	6120
GGTAATAATA CAGTATTTTT ATTAGCAAAT ATTTATGGTG TAGAGGCTAG CAAACCTAT	6180
ATATTATCGG ATTTAAAAAG GAAGTAAGAA A	6211

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7939 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCGGACTCCC CACGATTCTT CAAAATAACT GAGTATATTT CTATCTTGAT TTTCAGATAT	60
AAATTCTTCC TTCTGTGGCC TCTTCTTACG CTGAGAAGA GCTTCTCCGA CATGGCTTCT	120
TCCTTACTGA GCAAACCTT GAGCATAGAT AAGTTTGA CTGCAAGCGT CTCTGTATA	180
TTTGCTCCC TTCCACTAT TGTGGATAGC GAGGCGTCTT CTCATATCAG TCGTATAGCC	240
TATATAGTAG GATCCATCAC GAACTCCAG AACGTACATA TAAGCCTTAT GATCCATAAT	300
AAATCTCTTC GATTCGGGC GTATAAGAGC CATCATCATT GTGGACAATC AAAGGAGGTA	360

210

AGACCTTAAA GCCACTTGTT GAGCCATCCT TGATCGCCTC AATCAAAAGC ATATTGGCTT	420
CCTTTTCTCT TTTTGGATAA ACAAAGTGCA GCGGCTTAGG GGCTAGATTA TGTCGTTTTA	480
ACGTATCCAA AATATCCAGA AGTCGATCAG GACGATGAAC CATGGCCAAA CGCCCATTAG	540
ACTTGAGAAT ACTCTGGGCA CTACGACAGA TTTCTTCCAA ATTAGTCGTG ATTTCGTGTC	600
GAGCCAAGAG ATAATGTTCA CTCTCGTTCA GATTAGAATA AGGATTCACC TTGAAATAGG	660
GTGGATTACA CAAAATCATA TCCACCTTAC TCCCCTGAAT GTGAGCAGGC ATATTTTTC	720
AATCATCGCA GATGACCTGC ATTTGCTCCT CTAATCCATT CAAACGGACA GAGCGTTCAG	780
CCATATCCGC CAAACGCTCC TGAATCTCAA CAGACAATAT CTGTGCTTGA GTACGAGTGC	840
TAGCAAAAAG CCCCACTGCT CCATTCCCAG CACAGAAATC CACAATCAAC CCCTTCTTAG	900
GAAAACGTGG AAATCGTGAT AAGAGAACAC TATCCACCGA ATAGCTAAAA ACCTCTCTAT	960
TTTGAATGAT TTTGATATCT GTCGAAAAGA GCTGGTTAAT GCGCTCTCCT GATTTTAATA	1020
ATTGTTCTTC TTCCATGGTC CTATTATAGC AAATTCATAT TAACATTACA AAAAATATAA	1080
AACTCTAAAC TACTTCTTCT TTTTAAATG GTGCAGGGCT TCTCCAGTCC AGATPGGTAG	1140
CATTCGTCGA AAGGGAGCAA AGCCGTAGT AAAGCGGTCG CTTGAAAAGC GTCTCCGTCT	1200
AGGAACTGG TACTTTTCTT CCTCCAAAGT GCGGATAGAA AGACTGGCTT TCCCTGTAAA	1260
TTCATCTAAA TCCACTACCT GAACTTGAAC CTCTTCATCG ACTTTCAAGG TTTCATGAAT	1320
ATTTTCAATA AATCCTGTCC GAATCTCTGA AATGTGAATC AGCCCCGTAT CACCCGTCTC	1380
TAACTCAACA AAGGCACCGT AGGGCTGAAT CCCTGTAATA CGCCCCTTTA GCTTATCACC	1440
GATTTTCAATC TTAGTCTCG ATTTCAATAG TTTCAATTAC AACATCTTCA ACTGGCTTGT	1500
CCATAGCTCC TGTCTCAACA GCAGCAATGG CATCCAAGAC AGCGTAAGAT GCTTCATCAG	1560
CTAACTGACC AAAAACCGTG TGACGGCGGT CTAGGTGAGG TGTCCACCT TGATTGGCAT	1620
AGATTTCTGC AATCGGTCTT GGCCAACCAC CACGAGTAAT TTCTTTCTTA GAATAAGGTA	1680
GGTGTGGTT TTGCACGATA AAGAACTGGC TGCCGTTGGT ATTTGGACCA GCATTTGCCA	1740
TGGAAAGAGC ACCACGGATA TTGTAAAGCT CTCTGAGAA TTCATCCTCA AAAGATTGCG	1800
CGTAGATTGA CTCGCCACCC ATACCAAGT CAGTTGGGTC TCCACCTTGG ATCATAAAGT	1860
CCTTGATAAT ACGGTGGAAA ATGACACCAT CATAGTAGCC ATCTTTTGAA AGAGATACAA	1920
AGTTAGCCAC TGTTTATAGGA GCATGTTTCA GGAAAAGCTT GATACGTAAG TCTCCGTGAT	1980
TGGTCTTAAT AGTCGCAAGA GGACCTTCTA CTGTTTCAAT GTCTACTTGT GGAAAATGCA	2040
ATTCTTTTTC TACCATACCA AATACTTCTA AGGCAGCAAA AATGCCATCT TCTTCTAATG	2100
TTTTTGTAAT ATAATCTGCT TTTCTTTTGA TTTTATCATG AGAAATTCCT ATGGCAACGC	2160

211

TGATTCCAGC ATAATCAAAG AGTTCCAAGT CGTTGAGACC ATCTCCAAA ACCATGACCT	2220
TCTCTGGTTT CAAGCCAAGG TGTTCACAA CCTTTTCCAC CCCCCTCGCT TTGGAGCCTG	2280
AAATCGGCAC AATATCAGAC GAATGTTGAT GCCAACGAAC CATGCGAAGT TTGTCTGAGA	2340
GACTGTCAGG CAAGTGCAAG TCATCTCCCT TATCTTCAA AGTCCACATC TGATAGATAT	2400
CTTCTTTTTC ATGGAAATCG GGATCTACAT CTAAGTCGGG ATAAATTGGA TTGATAGCTT	2460
CACTCATCAT ATCGGTGCGA GTCGACAACT TGGCATCATG ACTCCCAACC AAGCCATACT	2520
CAATTCCCTC TTGCTTAGCC CAAGAGATAT ACTCCTCAAC ATCTGACTTT TCAATCTGAT	2580
GCTGATAAAT GACCTGACCT TTTTATCTT CGATATAAGC CCCATTCAAA GTTACAAAAA	2640
AGTCAGGCTT GAGATCACGA ATCTCTGGAA CAACACCAAA AATGCCACGT CCAGAGGCGA	2700
TTCTTGTTAA AATTCCCTTT TCACGCAACT GTTTAAAAAC AGTGGGAATT GTAGTTGGAA	2760
TAAACCCTGT CTTTGAATTC CGCAATGTAT CATCAATATC AAAAAAGACA ATCTTGATCT	2820
TCTTTGCCTT GTATCTTAAT TTCGCGTCCA TCTCACTACC TCTTTCAATC TAACTCTTTC	2880
CATTATATCA TAAAGTAGGC AAATCCCCTA TTTTCAAAA GTTTATCATT TTTATTTTAA	2940
TTCTTTGGAT GAGAAAAGAG ACATATTTAT GAAAAGCTC CATCGTGCTT TTAATGTGTT	3000
CTCTTGTTTT CAAACTCGTA AAAAGGGAGC CACTGATCCT AACTCGCTCT CTCATTTCAA	3060
AGCTTGTAAG AAAAGACCCG TTGGGGTCTT AATTCGCTTT CTTGTTTTCA AGCTCATGAA	3120
AAAGAGACCC AACTGGGTCT TTTCTTTAAT CTTCTTTTAC GAAAGGCATC AAAGCCATTA	3180
CGCGAGCGCG TTTGATAGCT GTTGTTACTT TACGTTGGTT TTTAGCTGAA GTTCCTGTTA	3240
CACGACGAGG AAGGATTTTC CCACGTTCTG AAACGAAACG GCTAAGAAGC TCAGTATCTT	3300
TGTAATCAAC ATATTCAATT TTGTTTGCTG CGATGTAATC AACTTTTTTA CGGCGTTTGA	3360
ATCCGCCACG ACGTTGTGTA GCCATGTTTT TTCTCCTTTA TAAGTTTAGT TGTCATTAG	3420
AATGGTAAAT CATCATCTGA AATATCCAAT GGGTTTGTG CTCCAAATGG ATTTTCATTA	3480
CGTGAAAAGT CTGGTACTGA ATTTGTAGGT GCTGAATAGT TTGCAGTTGG TGCAGAGTAA	3540
GCTCCACCTG TGTGACCCTC ACGCACACTA CGGCTTTCCA ACATTTGGAA ATTCTCAGCC	3600
ACGACCTCTG TCACGTAGAC ACGTTGTCCT TGCTGGTTAT CGTAACTACG AGTCTGGATA	3660
CGACCTGTCA CCCCATAAG TGAGCCTTTT TTAGCCCAGT TAGCAAGATT TTCAGCCTGT	3720
TGGCGCCACA TAACGACATT GATAAAATCA GCCTCACGTT CACCATTTTG ACTCTTAAAT	3780
GTACGGTTTA CTGCAAGAGT AAAAGTCGCA ACTGCTACAT TTGATGGGGT ATAACGCAAC	3840
TCAGCGTCAC GTGTCATACG CCCTACAAGT ACAACATTGT TAATCATAGT TTACCTTCTT	3900

212	
ACGCGTCAAT TTGACGATC ATGTGACGAA GAATGTCAGC GTTGATTTT GAAAGACGGT	3960
CAAACCTCTT AAGAGCTGCA TCGTCATTTG CTTCAACGTT AACGATGTGG TAAAGTCCTT	4020
CACGGAAATC TTGGATTTTC TATGCAAGAC GACGTTTTTC CCAAGTTTTT GATTCAACAA	4080
CAGTTGCACC GTTGTCAATC AAAATAGAGT CAAAACGTGC TACCAAAGCG TTTTTCAGCTT	4140
CTTCTTCAAT GTTTGGACGA ATGATATAAA GAATTTTCGT TTTAGCCATT GATATGTTCC	4200
TCCTTTTGGT CTAATGACCC CAAGACTTTG CAAGGGGTAA GTGAGGTTTC CTCACAATAA	4260
ACTATTATAC TAGAAAAAT TTTTTCACGC AAGTAAAAAC ACTAGAATTC GAAAAACGC	4320
CACATGGGCG TTTTCCTGTT CTTATGGTTT GATACGTCG AACATACGTG GGAATGGAAT	4380
AGCTTCACCG ATATGTTTTG TTCCTGCTGC GAAGGTACG ATACGTTTCA TACCGATACC	4440
AAATCCTCCG TGTGGAACGT TACCGTATTT ACGAAGGTCA AGGTAGAATT CATATTCTGT	4500
ACGATCCATG CCAAGTTCAT CCATCTTAGC GACAAGGGCA TCGTAATCTT CCTCACGCAT	4560
AGACCCACCG ATAATTTCTC CATAGCCTTC TGGAGCAAGC AAGTCTGCAC AAAGCACGCG	4620
CTCTGGATTT CCAGGAACGT GTTTCATGTA GAAGGCCTG ATGGCTGCTG GATAGTTTCA	4680
GACAAATGTT GGCACACCA AGTGGTTTGA AATCCAAGT TCGTGTGGTG ACCCAAAGTC	4740
ATCACCATGC TCAAGATGCT CGTAGTCAGC ATCTTCATCA TTTTCATGCT CTTGCAAGAG	4800
GTCAATGGCT TGATCGTAAG TGATACGTTT GAATGGCTCT GCAATGTAGC GTTTCAGAG	4860
TTCTGTATCA CGTTCCAAGG TTTCCAAGGC TTGAGGCGCG CGGTCAAGAA CACCTTGTAG	4920
AAGAGCTTTC ACATAAGCTT CTTGCAAGTC AAGCGACTCA TCATGTGTCA AGTATGAGTA	4980
CTCAGCATCC ATCATCCAGA ACTCAGTCAA GTGACGGCGT GTTTTGTGAT TTTCAGCAGC	5040
GAAACTGGA CCAAAGTCAA AGACACGACC AAGAGCCATA GCCCCTGCTT CTAGGTAAAG	5100
CTGACCTGAT TGGCTCAAGT AGGCTGGCGT TCCGAAGTAG TCAGTTTCAA AGAGTTCTGT	5160
AGAATCTTCT GCCGCATTTT CTGAAAGAAT TGGGCTGTCA AACTTCATAA AACCGTCTT	5220
GTCAAAGAAC TCATAAGTTG CATAGATAAT AGCGTTACGG ATTTGCAACA CAGCTACTTG	5280
CTTACGAGAG CGTAGCCACA AGTGACGGT ATCCATCAAA AAGTCTGTTC CGTGTCTTT	5340
TGGTGTGATT GGGTAGTCTT GAGATTCACC GATCACTTCG ATGTCTGTGA TGTCCAACCT	5400
ATAGCCAAAT TTAGAAGCTT CGTCCTCTTT GACAATACCT GTCACATAAA CAGACGTTTC	5460
TTGGCTCAAG CGTTTGATAA CATCAAACTT CTCAAGTCCC ACTTCTTCAC CAAATTTTTC	5520
GACAAAGTTT GGTTTAAAAG CCACACCTTG AAAGAAGGCT GTTCCATCAC GCAATTGTAA	5580
GAAAGCGATT TTTCTTTTTC CTGATTTGTT GGCAACCCAA GCGCCAATCG TCACTTCTTG	5640
ACCAACATAG TCTTTTACGT CAATAATCGT TACACGTTTT GTCATTATTT TTCCTTTTCT	5700

TTTTTATTCT TTATGGCAAA CCACCTCTAT ATTGTTCCCA TCCAGGTCAA TCATAAAAGC	5760
AGCATAGTAA ATCGGATGCT CACTTCGATA ACCAGGAGCC CCATTGTCTC GCCCACCTGC	5820
CTCTAAGCCA GCCTCATAAC AAGCCTGAAC TTCTTCCTTA TTTTCTGCTA AAAAAAGCAA	5880
ATGAACAGGA TCTTGTGTTT CCTGAGTCAG CCAAAAATCA CCACCAGGAT GAGGGCTGTT	5940
CGGGGATAGA AACTAATTA GAGAACTAGT CTTAAAAGCC AATTTATAGT CCAAAGGAGC	6000
GAGAAAACCTC CTATAAAATC CTTATGAAAT TTGTAAATCC TTTACCTTAA TCTCAAAATG	6060
ATCAATCATT CTCACCTACC ATAAATGCTT TCAAGCGTTC GACTGCTTCT TTAAGCGTGT	6120
CTAGGTCTGT CGCATAGCTG AGGCGGACAT TTTCTGGTGC TCCAAATCCA GCTCCTGTTA	6180
CCAAGGCCAC TTCGGCTTCT TCTAAGATAA CAGTTGTAAA GTCTGTCACA TCCGTGTAGC	6240
CTTTCATCTC CATGGCCTTT TTGACATTTG GGAAGAGATA GAAGGCCCTT TGCGGTTTGA	6300
CCACTTCAAA TCCTGGTACC TCTGCAAGGA GGGGATAGAT GGTATTAAGA CGTTCCTCAA	6360
AGGCCTGACG CATGCTTTCT ACAGTATCTT GCTCACCTGA TAGAGCCTCA ACTGCTGCAT	6420
ATTGGGCTAC TGCTGACGGA TTCGAAGTTG TTTGACCTGC AATCTTGGAC ATGGCAGCGA	6480
TAATGTCTGC TTCTCCAACG GCATAACCA TCCGCCAACC AGTCATGGCA TAAGTTTGTAG	6540
ACACACCATT GATGACCACT GTTGTCTTGC GAATCGCTTC CGATAGGCTA GAAATCGGTG	6600
TGAACTCATG ACCATTATAA ACCAAGCGGC CATAGATATC GTCTGCTAGG ATGAGAATAT	6660
CATTTTCTAC AGCCCAGTTT CCAATTGCCA AGAGTTCCTC ACGGGTGTA ATCATACCTG	6720
TGGGATTAGA TGGCGAATTC AGCACCAAAA CCTTGGTCTT GTCAGTGCGA GCTGCTTCTA	6780
ACTGCTCTAC GGTACCTTA AAGTGATTGT CTTCTTAGC AGAAACAAAG ACGGGAACGC	6840
CTTCTGCCAT CTTGACCTGA TCTCCATAGC TAACCCAGTA TGGGGTTGGG ATGATGACTT	6900
CATCACCTGG ATTGACCACA GCCATAAGA AGGTATAGAG AGAATATTTG GCTCCCGCAG	6960
CGACTGTCAC TTGATTTGAC GCTACAGAAT AGCCGTAAAA GCGCTCAAAG TAGCTATTGA	7020
CCGCCGCCCTT AAGCTCTGGC AGACCTGAGG TTACTGTATA AAAAGAAGCA CGCCCATCTC	7080
GAATCGATGC AATGGCGGCA TCTTGATAT TTTTGGGAGT AGTGAAATCT GGCTCACCCA	7140
AGGTTAGAGA CAAAATATCT CTACCCTCAG CCTTCAGTGC TTTGGCACGG GCTCCAGCAG	7200
CCAAAGTCAC ACTTTCTTCC ATTTCTAAAA CACGGTTGGA TAGTTTCATA GGCCCTCCTT	7260
GTTGACCAAT GCTCCTGTTT CAAAATCTAC TAGATAAAAA TCAGATCCTG ACTTAACTTC	7320
CCAGATTGGC TTATCTTGAT AACGGCCAAA GGTATCTTG TCAATCTCGC CAGCTCCCTT	7380
TTCTTAGAA ACCGTTTCTG CTTTCTCTTG TGAAACACCC TGATTTAGCT GATAAACGTA	7440

214

AATCTTATGG TCATCTTTAC CAATCAGGAC AGCAAGCGCT TCTTGCTGTT TGTACGACC	7500
AAGAACGCTG TAATAAGATT CCAAGCCATT GTATAAATCA ACCTGATCAG CCTGCTCTAA	7560
TCCTGCATAC TGCTGAGCTA ATTTTCTCC TTCACTTTTA GCTGTTTGAT AGGGTTTCAT	7620
GCTAAGAGAA ACCATATACA GAAAGGAACC ACTGATAACC ACAAACAAA TCGTCATCCC	7680
TAGACCATAC TGCCACAGTA GATTATTTTT TGCTTTGTTT TGTCTTTTTT TCACTCGTCT	7740
ATTTTACCAT CTATTAAGCT TTATTACAAG TGAATATAAG AATACTCTTC GAAAACTCT	7800
TCAAACACG TCAGCTTTAT CTGCAGACCT CAAAGCTGTG CTTTGAGCAA CCAATTCTAT	7860
TTCTCCCTTC AAACAAAACC GATTTTGAAA GTGAAACAGT TCTTACTTTT TCAGTCACAA	7920
ATGATTAGAG TTTGCCGGG	7939

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9897 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCGCTCTACC GTCAAATAAT TACCATTTTG TTAAATACCG AAATTTTAT CTA CTGAAAA	60
TTCAAGTGGT CTGTTGGTAC GATCGTCGTA TACAGTACCA TTCTCAGGAA TAGTATAATT	120
GTAATCAGTA TCACCTTGTT TCCTTAATTT AAGGTAATAA TTACCATCAA TTTGTTTATA	180
ACCTGAATCT TTTCTAGTTG CTTCTCTAAA ACTTACTCCA GCAGGCATCA CATCAGCAAA	240
CATGAGTACT TGTTTGTTCT TTTTTC AAC AATAACAGAG TCAATATAGG TTGCACCACC	300
GCTGATTTGT AAGTCACGTC CACCAACTTC ACGAGGCCAT TCTAATGGTA CTGGCGCAAA	360
ATCATCGAAT GCCAATGTTA ATTTTGTTT AGTCCATGTC TTACCATTAT CATCACTATA	420
ACTTGTAGCA ATATTAATTT TATTCAAGAA ATCATGAGTT CCACCGTAAC GAGCGTCAAT	480
GCTTGAAAAT ACCCGACCAT TGCTAAAAGT ATACAGAACT GGAATACGGA AATAGTTAGA	540
ACCTGTTGTA TCATTAGCCG TATAAATTAA ATGTCCAGTA ACAGCGTTTG TTGTCATCTT	600
TTTAACAGTT TCTTCATCCA ATGCACTATT AAAGAATTG ATATTTTCTA GTGTCCGTT	660
AAAACCAAAAC GCCGTTTTTC CTGCACGTTT CACTCCCCCA AGCATATAGT AATCAATACC	720
TTTAATATCC TTGATGTTTA GGAAATTATC CACTTCTTT TCTACTACTT TTGTACCATT	780
TGCGTATAAA GAATATGTTT TTTTGACTGA ATCTGCTACT ACTGCAACAG TGTTAGTCAC	840
AGCCTCTTGT TTGTACTTAC CCCAACTGA AGCAGGTCTG GATACTAGGT TATTTTATTT	900

215

GGAAGAAGTA TCACGCGCTT CCATCCCCAA CTCACCATTG TCTCTAAGGA ACACATCTAC	960
ATAACTATTT TGTGACCGG GTTTGGAATT AGATATCCA AACAGAGCTT GTAAGCCTTT	1020
CTCACTTGAC TGATTGTACT TAATCACTAC AGTAAAGTCA CCGCTAGTAA ATTTATCCTT	1080
TAACCTTTTA GTAACATTTT CTCCGCCCCC TGTAAAGTA ACATTATTTT TTTCTAAGAC	1140
AGGAGTTTCT TCCGCTGTAG AAGATGGATC CTAAACAGTA GTTCAACTG TTCGAGGTTG	1200
TACAGTAACT TCCGAAGAGT TATCCGATGT AGGTTGTACT TCCGAAATCG GAGTCGTTGG	1260
TGCAACAGGT TGACCAACT TTGGTGTGA TACTTCAGAA GTTTCAGTCT CCTGAGCTGC	1320
AACTGAGTTA GCAACAAATG CTGATAATAC CACTACAGTA CCTAAGGTTA CATATTGTTT	1380
AATATTTTTT TTCATTTTAT TTTTCCTCGT TTAACACTTT GATAACAAAGT TTTTAAACAG	1440
TTTCATCATT GCAATGAATC TTTGGTTGGT GAAGATCTTC TTCAAAAGTC ACCAACATAT	1500
TCCCTGGAAG CAATTCAACA ATTTGATAGT CTTTGCTATC GTAAAAAGCA ATATCCTTCT	1560
CTTCGCTAAA AGGTACACGT GACTGGGCAC GAACTGGGGA AGTTACTGCC ATTTTTCAG	1620
TATTTTCAAC AACAATATGA ATATCTAAAT ATTTCTTATG AGTTTCAAAA ATATCTCTCG	1680
GAACTCCATC AGCTAGATAA GTCATACAAT TTGCAAAAAC ATTTTCCCCG TCAATATCAA	1740
TTTTTCCATC AACTAAATCT GTCAAATTTG TATTTTCTAA AAAATCACAG ACTTTTGAAA	1800
AATATTTTATT GACAGAAGCA TATCGTTTAA AATCAGATTG TTCAGAAATA ATCATATTAT	1860
TTTCTCTTTT CTATTAGTGA CGAACTTCCC AACTTGAATC CGCTTTAATT TCTGTAATAT	1920
CATGAATCGT TGTATATTTA GGTGCAGATA CTTTATTTCC AGTAAGAACA GATACAATAT	1980
AACCTGAAAC TACTGATACA GAGATTGAAA TCAATGAATA TGCCCAGTAG CTAACAGCTG	2040
TTGGAGGAAG GAAGTATTTA ATAAATACCA TGACGATGGT TGATACAATC AGCGCTGCAT	2100
AAGCACCTTG TTTATTTGCT TTTTGTAGAA CAAATCCAAG AATAAATACA CCACCAAGTA	2160
GACCAAGTAC AAGTCCCATG AACTATTGA ACCATTCTGA TGCAGATTTA ATATCTGAGT	2220
GAGCCATGAC AATGGAACA CCAATTGAGA ATAAACCTAC TGCTAGAGAT ACGAATTGTG	2280
CAATTTTCGT ACGACGATTG TCTGACATAT TTTTAGAAAT GACATCTTGA ATATCCAATG	2340
TCCATGAAGT TGCAACAGAG TTCAAACCTG TTGAAATAGT TGATTGAGAT GCTGCATAAA	2400
TCGCTGCCAA GATCAAACCT GTGATACCTA CTGGTAACTG GTATGCAATA AAGTACATAA	2460
AGATTTGGTC TTGAGGGATA TTGCTAGCTG CACTATCTGC ATTTTGTACT TGATAGAATA	2520
CGTACAAGCC TGTACCAATC AAGTAAAGA CTGTTGCAGT TGCAAGTGAC AAAACACCGT	2580
TTGTGAACAA CATCTTATTA AGTTTCTTAA TATTTTGTGT TGTAGTAAAA CGTTGAACCA	2640

216

AATCTTGAGA TGAAGCATAG GAAGACAAGA TTGTAAAGCC TGAACCCATC ACAATTAAAA 2700
AGATGGAGTT TGAAAGCAAG TTAGGATCGA AAAGTTTTTC ATTTGCAGCA AGGAATTTCC 2760
CGTTTGCTAA TGTTCCTGCT ACTGCACCAA AGCCACCTTT AATATTAGCA ATCAGTACAA 2820
ATAAAGCTAA AACGACACCA CTAATCAGAA TCACACCTTG AATAAAGTCT GTCCATAATA 2880
CGGATTTTAG ACCACCAGTA TAAGAATAAA CAATTGCAAC TACACCCATC AAAATAATCA 2940
AAATATTGAT GTCAATTCCT GTCAATACTG ATAAACCAGC TGATGGGAGG TACATAATGA 3000
TAGACATACG TCCCAATTGA TAAATAATAA ACAAGAGTGC TGAAATAATA CGAAGTGCTT 3060
TAGAATTAAA ACGTTTATCC AAGTAATCAT ATGCCGTATC GATGCTATC CGTGCAAAGA 3120
TAGGTAAGAT AAAACGAATT GTCAGTGGAA TAGCTACTAC CATCCCTAAT TGAGCAAACC 3180
ATAAAATCCA GCTACCTGCA TAAGAGCTAC CAGCGAGTCC CAAGAAGGAA ATCGGACTGA 3240
GCATTGTGGC AAAAATGGAT ACCGAAGTAA CATACCAAGG AACCGAACCA TCTCCTTTAA 3300
AGAACTCTTT TCCTTTCATC TCTTTTTTAG AGAAATAGAT ACCTGCAACC AACACCGCAA 3360
GTAAATAAAC AATCAAGATA ATTAAGTCAA TTATTGTAAA TCCTGTTGTG CCCATAACAT 3420
ATCTCCATAT TGATTTTATT TATTATAAAA ATTCTTTTCG TGCTTGTTGA ATAAGTCTG 3480
CTGCTTGTTT TGCAACTTCC AAGTCACCTT CTGCCAATGC TTCTAAAGGT TGACGAACAG 3540
AACCTAAATC AAGTTTTTCA TTTAGACGCA AACTTCTTT TGCTACAGCA TACATATTTG 3600
CCTTACCTGA TATCATCTTA TAGATAACTT CATTGATAGC ATATTGAAGT TTTTTCAGTG 3660
TATCTAAATC TCGTTCCTGA ATCAAATTT CCAATTTCAA GAACAAATCT GGCATAACGC 3720
CATAAGTACC ACCAATACCA GCTTCTGCTC CCATCAAGCG ACCACCAAGA TATTGTTTAT 3780
CTGGACCATT GAATACAATG TAATCTTCTC CACCTGCAGC TACAAACATT TGAATATCTT 3840
GTACAGGCAT AGAAGAATTT TTAATCCAA TCACACGAGG ATTTTGACGC ATTGTTGCAT 3900
ACAAACTACC AGTCAACGCA ACCCCTGCCA ATTGTGGAAT ATTATAGATA ATAAATCTG 3960
TATTTGACGC AGCTTCACTC ATTGCATTCC AATATGCTGC GATTGAATAC TCTGGCAATT 4020
TGAAATAAAT AGGTGGGATA GCTGCAATAG CATCGACTCC AACACTTTCT GAATGTTTTG 4080
CCAATTCGAT ACTATCTTTC GTGTTATTAC ATGCAATATG GTTGATAACT GTTAATTTAC 4140
CTTTAGCAAC TTCCATAACA GCTTCAATAA TTTGTTTACG ATCTTCTACA CTTTGGTAAA 4200
TACATTCACC TGAAGAACCA TTTACATAGA TACCTTTTAC ACCTTTGTCA ATGAAATATT 4260
GTACCAGAGA TTTTACACGA TCTTGGCTAA TTTTACCATT TTCATCATAG CAAGCATAAA 4320
ATGCAGGGAT AACGCCTTTG TATTTAGTTA AATCTTTCAT CAGATTTCTC CTTTATATTG 4380
TTTTTTATTT GATGACATTA ATAAATCGCT GAGCAATTC TTTTGGACGT GTAATCGCTC 4440

217

CACCAATGAC TACACTGGTA ACACCTAAAC TATAAGCTTT TTTTAATTGT TCTGGATAAT	4500
GAATTTTTCt TCGGCAATTA CCGGAATATT AAAATCAGCC AATTTTTCa TTAGTTCAAA	4560
ATCAGGCTCA TCTGATTGTA CACTTGTA CTGTGTAACCT GATAATGTTG TACCAACAAA	4620
ATCAACGCCT GATTAAATG CATAGAGACC TTCATCTAAA TTACTTACAT CCGCCATCAG	4680
CAATTGATTC GGATATTTT CTTTATTTT TTTGATAAAT TCACTGACAA CTAAGCCATC	4740
ATATCTTGGT CTTAAAGTTG CATCAAATGC AATGACTGTT GTTCCGCATT CTACAAGTTC	4800
ATCTACTTCT TTCATCGTAG CAGTAATATA TGGTCTTGA GGTGGATAAT CCCTTTTGAT	4860
AATTCCAATT ATTGGTAAAT CTACTACTT CTGAATTGCT TTAATATCAC GCACAGAATT	4920
TGCGCGAATG CCCACTGCTC CTGCCTCTAA AGCTGCTTTA GCCATAAAAG GCATCAAGCT	4980
AAATTCCTCA TTATAAAGG CTTCAACCAGG TAAAGCTTGA CAAGAAACAA TGACTCCACC	5040
TTGAACTTGG CTATAAATT TTTCTTTAGT CCAAATTTGG CTCATTTTAT TATTCCTCCT	5100
TATGGATAAT AGTTTGATTG TAATAATATT GTCTCTCTGG ACTTCCAGA TAATTAGAGA	5160
ATAAGCAGTC TGTAATTAAA AGTATTGGAA ACTGAGGTGA TATGCGATTG CCATACGAGA	5220
GATGATCGGT CGAAGCTAAT AACAATAGTT CATCAAAGAA ACAATCTTCT TCGTCAAATT	5280
TTCTTGTAAGT CATTAAACT GTTTAGCGC CTTATCTGC AGCTTTTGT AGACCTTCTA	5340
GTACAATATC AGTTTGACCT GAAATGGATG CTCCAATGAC AAGGCAATTT TCATTAAGTA	5400
GTAAGCTACT CCACAAAATC ATATCCTCGT CTGATAATAC TTCACCAATC ACTCCGAGAC	5460
GCATAAATCT CATCTTCATT TCTTGTAAG CAAGAACAGA ACTTCCTTTA CCGTAGAGAT	5520
ATACACGCTC AGCAGTTTCT ATCATCTCAG CAATACGCTC AAGTTGAACT TCATCAAGAA	5580
CCGTGTAAGT TTTTCTCAAC ATTTCTCAT AGTCGGATAA AACTTTTCT GTGCCTCTG	5640
TATATAATGC CAACTTTTCT TTCTCATGAA TCATCTCTG GTATTTGAAA ATGAATTGTC	5700
TAAAACCTTT AAAACCACAT TTTTCGCAA ATCGAGTCAA TGTGCTTTG GATACATTAA	5760
GGTATTCGCA CAATGCTTTA GATGAATAAT CATTGAGAGG TTGCTGTTT AAGAAGAATT	5820
TAGCAATGTC TTTTCAGCA TATGCCATAT TTGGTAAGTT AGCTTCTATC ATTGGAATTA	5880
GTTCTTTTGG CAGTAACATA TGAGCTCCTT AGTTGAAGTA AACGTTTACA TTCTTTATTT	5940
TAACACTTTT TTTTTCATC AATATTTTC ATAAATTAGA AACTAGTTTC CAATTTCTTT	6000
CGTTTCATAA CAGAACAACA AACATAAAAA TATAATAGTT TTTATTCCTT TTATCGTAAT	6060
TATATGTATT GTAAGAACGT TTATCACTAA TAATATGTTT ATATTAAAT ATTTTAGTAA	6120
TATTTTATTT TGGTTTATT ATTTCTTTC GGAATTTCTA TATAATATT TATTTCTAAA	6180

218

AAAATTGAAA AAATATTTCT AGTTTCTTTA TTTTATATAG GTAATATATT TTATTTCTAA	6240
ATTAAAAGAG AATCCATAA AACTACAGA TTTATGAGAT AAATCAGGTC ACCTATTTTA	6300
AAAAAGCAGC AACTATAAA CTAAAAAGTT CCACACCAAA TGTAACCCCA TACTTCCCCA	6360
TAAGTCAGAT TTATAGCGCA CCATACCTAA AAACATTCCA AGTGAAACGT ACAGACACCA	6420
AGCTAGAATG GTTCCTGGAT GATGTACTAA GGCAAATAAA ACACTTGTCA AAGCAACTCG	6480
AATATCTAAT TTTCTAACCA AGTTCCATAA AATTTCACGA TACAGAAATT CTTCAACCAT	6540
ACTCGCATTG ATTAAGAACA ATAAAAATGA AAACCAAGGA ACTTGATGTT GAAGGCCAAT	6600
TAAATTTGTT TGATTCGTGC TTCCTTGAGC ATGAATCAGG CTAAACATA GACTTATAAT	6660
CAGTAGACTA GCTAGTCCAA TACCAAGGCA TTTCATCCTA GTTTTCATAT TGACCTTGAC	6720
CACTTGTTTT CGTTGACCAT ACATCCATAA AAAAGAAAAA AGAGACGCAC CATAGAGAAC	6780
CTGTAGTATA GTTAACTCAC CGATACAAAG AAATTTCAAT AAGTATAGAG ATACCAATAG	6840
GACATTTACT TGTGGAATA TATAAACTGG AATTATCTTT TTCATAGTTA CCTCCGAAAT	6900
AAATCTTCAT AATCTAAATC TAATATCTGC ACAATCCTTT CTACCCATGG ACTTTGAGGC	6960
ATTCGTTGTT CCATCTTGTA GTGGCGAATC TTTTGATATA AACGATTCAA TTCACTTGGA	7020
TAGTGAAACT CTCCCGCAA CATTTTCTG GTTAACTCAA TCCAGCTGAT ATTTCTTTCA	7080
GCCAAAATAA TGGACAAGTT CTCCCAAAT CGTTCAGCCA TATTCTTCT CTTTAGTTA	7140
GATAAATAAT GTGTTGyGC CATGTAAATC AATTGTTTCG TATCTCTTGG CAATAGAGCT	7200
CTAGCCTCTT CCAAATTCAG ACTTGGATAA ACCCGCTTAT TTGAAACCAC AAAAGGAAGT	7260
CCGATGGTTA GTTCAGGATT TTTTAAAAT ATCTCAACGA AATCCGTAA TCTTAGATTG	7320
TCACGGTTCT TAAATCGTAA TAAATTGGGA GATAAAAACT CAAAACAATC TGAAGAATAG	7380
CTCATCATCT CAATTAATTT GTCCTTTGTC ATTTCAGAAA CTGAATGACA AGATACCTCA	7440
ATGCCATAGT TTTGGAAGAA GTCTAAAAGA AGTTGATTTT TTTGGCTATT TTTACTTAGA	7500
TAGAGATCAA TCATGGGAGA CCTCCAACAA ATTTGCTTCC ATTTGATATT CTGAGACGAT	7560
TAAGGAATCT AACAACTTTG AGAAGTTAAT CGATTCTTG TCTTCATCAT AAGCTTTTAC	7620
AGTTACTTGG GTTGTAAGTA TCCCCTCTTT TCCCTCGGCT CGATAGTCTT GTCAATATAA	7680
AACAAAAACA AGATTCTGAT TATCATCTAC AAAGGCATTA ACTCCGTTCT TTATATCCTG	7740
ACTTTCAAGG AATTCCATAA CGTTTGAAG ATAGGATTCA TAAAATAGTG GGTAATTATG	7800
TTTTTTATGG TAATCATCTA AAAATGTTAC CTCAAACTCA CATGGATAAT TGGGCATCAA	7860
AAATATTTGT TCATCCAGCT GTTTGATTTT TGCATCATGT AATTCTGTTT CTAATTCATC	7920
ACAATCTAGT ATTGATTCTT TATTTAATGC TTTTATCTTT TCCTCTATT TCTTTAATT	7980

219

TCTTTGCGAT TGCGGCAATC ACAGGAACGG TTACACTATT ACCAACTTGT TTATAGAGCT	8040
GACTATTAAT AGAGACTTTT CTAGCAGCTT CAAAAGCCTA ATCAGGAAAG CCATGCAATC	8100
GAAAACACTC TTTAGGAGTG ATTCGTCGTA TTCTCAAACG GTAAAATTGT CCATCTATTA	8160
AAACACCAGC TACTTGGTAA ACTTGTTTAT CTTCTCCTTC ATAGCTAGCC ACTACTACTC	8220
CCATTTGACC ACTAGTTGTT AACGTATTAG CTATACCTTT TCCAACTCTA CCACGACGAT	8280
ACTGAGAACT TGGTCTTTCT AAATTGATTG AATCCCCAAT CTCTGCTTGA GCATATCCTT	8340
TTTTCGTTGC TTCCCGTACT TTTAGAAATT GGATTGGTTC TGGAATTAGT ATTTTGGGGA	8400
TTTATCTCC TCCTTGCATC GTAGTCAGTG TTGGAGATAA GCCCTCACTT CCATAGACAC	8460
GACCTGTCTC CTTAAAGCTA GTCGGTAAAT CTCCAACAAC GACAATGCCA TAACGATCCT	8520
GAGTATTTAA AGTAAACATC GGCTCTMGAT TTTCTTAAA GCGTCTCCCA TTTTGTCTCT	8580
TGTCTAATCT ATCTGGTCTC ATACAAGGAA TCGCAACTTT AAATCCTTCT CCTTTACCAC	8640
GAACTAAGGT TGGCGCAAGA CCTTCTGAAT AATAGACTTT ACCGCTCATT CCACCTCTTG	8700
ATGGATTCAA ATTTCTAGT GCTTTCAAAG TCTCAGAGTT AGTTGCTTGA CCTTCTCGTC	8760
TGAAAGGAAA TAAGAGCTG GTACCTTTCT TTCTAGAATG TCCGATAATA AACACCCTCT	8820
CTCTGTTTTT GGAACGCCA AAATCCTTAC TGTTAAGCAC CTGCCACTCA ACATCAAACC	8880
CCAATCATC AAGTGTGTA AGTATTGTGG TGAACGTCCG TCCCTTATCG TGATTGAGTA	8940
GGCCTTTAAC ATTTTCAAGA AAAAGAAAAC GTGGTTGGAT TTGTTTGGCC GCGGAGCAA	9000
TTTCAAAGAA CAAAGTTCCT CTAGTATCTT CAAATCCCAA TCGTCTTCCT GCGATTGAAA	9060
ATGCTTGACA AGGGAATCCC CCACAGATGA CATCGACTTT CCCTCTAAGT TTTTAAATT	9120
CGTCATCTGA AACATCTCGT ATGTCATGAA ATTCTATTTT TCCTTCCGTT TGAAAAATGG	9180
ACTTATAAGA TTTCCTAGCA AATTTATCAA TCTCACAAA TCCCAAGCAC TCATGCCCTT	9240
GAGCTCCAT TCCCATCCTA AAGCCTCCTA TCCCAGCAA TAAATCTAAA ACCCAAATCA	9300
TTACATCTC TCTCAACTAG ATGTAACCTA CAAAACCCCT GACCTCATGA GCCACTTTCT	9360
TCCTCCTCAT GAGGTCAGTT TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT	9420
TTCTCAAAA GGCAGACTC CTCCCTTGGT TCGTCACAG ATTTTTCAT CTCGACTGTT	9480
CTTTAATGCA TCATTAACGA CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC	9540
AGGTTGACTT TTCTAATCCT AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA	9600
CTAGACAATT TGAGGAGCTG CTTGCGTCCT GTTCGAACAC ATTTTCCTAC CACGTGAAGA	9660
AAAAGATGGC GGAAGCGTTT GATTGTTAAA GTTTGGAAGT CACCTCCAGC TAGATGTTTG	9720

220
AGAAAAAGAT AGAGATTGTA GCGGATACAG CTCATCATCA TACGAACTCG TTTTGTGATTA 9780
AGGTTGAACT ATCCGTTTTA TCGCCAAAAA ATCCCTCCTT CATCTCCTTG ATGAAATTCT 9840
CGGCTTGACC ACGTCCACGA TAAAGCTGAA ACTGGTCTTG GCTTGTTCGG GTACCGA 9897

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8148 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCGTGGAACA AGCCAAGACC AGTTTCAGCT TTATCGTGGA CGTGGTCAAG CCGAGAATTT 60
CATCAAGGAG ATGAAGGAGG GATTTTTTGG CGATAAACG GATAGTTCAA CCTTAATCAA 120
AAACGAAGTT CGTATGATGA TGAGCTGTAT CGCCTACAAT CTCTATCTTT TTCTCAAACA 180
TCTAGCTGGA GGTGACTTCC AAACTTTAAC AATCAAACGC TTCCGCCATC TTTTCTTCA 240
CGTGGTAGGA AAATGTGTTT GAACAGGACG CAAGCAGCTC CTCAAATTGT CTAGTCTCTA 300
TGCCCTATTCC GAATTGTTTT CAGCACTTTA TTCTAGGATT AGAAAAGTCA ACCTGAATCT 360
TCCTGTTTCT TATGAACCAC CTAGAAGAAA AGCGTCGTGA ATGATGCATT AAAGAACAGT 420
CGAGATGAAA AAATCGTGTG ACGAACCAAG GGAGGAGTCT GCCCTTTTGA GGAAATCTAG 480
CGAGGAAAAA CGATACTGGA ACAGCAGAAA GTAAACTGA CCTCATGAGG AGGAAGAAAAG 540
TGGCTCATGA GGTGAGGGGT TTTGTAAGTT ACATCTAGTT GAGAGAGGTA TGAATGATTT 600
GGGTAAATAC AATGAGCTTG AAAGAAGTAG CAAACTCACC AAGCGCCAAT TCTTTGAGAA 660
TCAGATGCTG GATTATACCA TCATTGCGCA TGAGAGTTTT GAAATCATCC GTCATTCTGT 720
CTACCAGACA GATGATCGTG AAGTGGAAAA TGCTCTGGCT TTTGAAGTGA AAAATGATGA 780
AACAGACAAG CTGATTCTGT TATTAAGCGA GGATATTGGT GTAGGTGAAA AATTGTGCCT 840
CGTTGACGGA ACAAATGTC GTGGAAAATG TTTAGTATAT GATAAAATAA ATGAGAGAAT 900
GATTCGCTTG CAGTGCTAGA AATAGGCATT TTGAATAGTG AATATGTTAT AATAAGTATT 960
AGTAGGAGGT GTTTTAGATT GGAGAAGAAA CTGACCATAA AAGACATTGC GGAAATGGCT 1020
CAGACCTCGA AAACAACCGT GTCATTTTAC CTAAACGGGA AATATGAAAA AATGTCCCAA 1080
GAGACACGTG AAAAGATTGA AAAAGTTATT CATGAAACAA ATTACAAACC GAGCATTGTT 1140
GCGCGTAGCT TAAACTCCAA ACGAACAAAA TTAATCGGTG TTTTGATTGG TGATATTACC 1200
AACAGTTTCT CAAACCAAAT TGTTAAGGGA ATTGAGGATA TCGCCAGCCA GAATGGCTAC 1260

221

CAGGTAATGA TAGGAAATAG TAATTACAGC CAAGAGAGTG AGGACCGGTA TATTGAAAGC 1320
ATGCTTCTCT TGGGAGTAGA CGGCTTTATT ATTCAGCCGA CCTCTAATTT CCGAAAATAT 1380
TCTCGTATCA TCGATGAGAA AAAGAAGAAA ATGGTCTTTT TTGATAGTCA GCTCTATGAA 1440
CACCGGACTA GCTGGGTAA AACCAATAAC TATGATGCCG TTTATGACAT GACCCAGTCC 1500
TGTATCGAAA AAGGTTATGA ACATTTTCTC TTGATTACAG CGGATACGAG TCGTTTGAGT 1560
ACTCGGATTG AGCGGGCAAG TGGTTTTGTG GATGCTTTAA CAGATGCTAA TATGCGTCAC 1620
GCCAGTCTAA CCATTGAAGA TAAGCATACG AATTGGAAC AAATTAAGGA ATTTTACAA 1680
AAAGAAATCG ATCCCGATGA AAAAATCTG GTATTTATCC CTAAGTGTG GGCCCTACCT 1740
CTAGTCTTTA CCGTTATCAA AGAGTTGAAT TATAACTTGC CACAAGTTGG GTTGATTGGT 1800
TTTGACAATA CGGAGTGGAC TTGCTTTTCT TCTCCAAGTG TTTGACGCT GGTTCAGCCC 1860
TCCTTTGAGG AAGGACAACA GGCTACAAAG ATTTTGATTG ACCAGATTGA AGGTCGCAAT 1920
CAAGAAGAAA GGCAACAAGT CTTGGATTGT AGTGTGAATT GGAAAGAGTC GACTTTCTAA 1980
AATGAAGGAA AATGACTTGC AATCTCTGTT AAGAAATAAA ATAATCCAC CTAGAACAAG 2040
CTAGGTGGGA TTATTTGCCT ATGAAATGAG AAATTATGGG AGCAAGCTCC TAAATCAACT 2100
GTTTTTGATC TACTTCTTTA ACTACTTGAT AAAAGTTATA GAAGTAGGCC AAACCTGAAA 2160
TGATGGTTAC GACTAGGAAT ATTGAAAATT TCCATTGGAC AGGGTTGGTT AAAAGTTGTG 2220
GAAAGGATAT GAGGAGAAAG AAGAGGGCTG CGTTGAGGAC AGGTATCCGT TTTGATTGTA 2280
TTTTCTCAAG TCCTTTATTG AGCGCAGGAA GAAAGAGGAG TAGGAGTAGT AAAACTGTAT 2340
GAGAAATAGC TCCTGAAGTA AGGGCGAAGA AAAGGAAAAT ACTGATAAAA ACATGAATGA 2400
TCAGTAGTCT AGCTAGTGAT TTCATAAGGC ACCTCCTAAT CCTGGTCTTT TTTAGCTCTT 2460
GCAATACGAA GTGAGTCGAC AATATGTATC ATCACTCCGA AAAAGAAAGC TCCCAGTATA 2520
GTTTTAAAAA TATGTTTTGT ATTTAGAAGA GAACTGATAA AATTGGAAT TCACTTGTT 2580
AGGGTATCAA TGAGTGGAAT TATAAAAAAT ATCACTGTTC CATAAATCGA ACCTGCTTTC 2640
AGACCAGGAT AACGTAACGT TTTCTTTTCT TTTTTCATGA GTTTCCTCCT AATCCTCATC 2700
TTGATTTTTC TTAGTTTTTG CAATGCGACG GGAGATGAGG AACTGTATGC TCGCTCCGAA 2760
GAAAATAGAA CCGAGAATAC TTGATACACC ATTTCTTATA GTGAGAAGAG AATGAAAATA 2820
GTCCTGACCT TCATCTATGA GTATCCTGAG AAGAGGAGTT ATAAAAACA TCCATAGACC 2880
AAAGAACAAA CCTGCTTTCA GACCTGGGTA GTGTAGTTGC TTGCTTTCTT TCTCATTCAG 2940
CATATCTGGT TCAATGACTG TGATGCCTGT TTTTTCATT TGGTAGGTGA CATAGCCAGA 3000

222

AGCGATGAGG GCAATCACTA AAATCAGAGG AGGATAGATT AGAGCCACTT CTTGAGGGTA	3060
TTTATAGGCC AGAAGGAGTG GAATAAGATT TCCGAAAATC ATCAGATAAA AGAGGATGAT	3120
AAAGACTTGG TTCCCAATAC TATCGGCCTC ACGCCGTTTG TATTCGTCAA GGGGACCAGA	3180
AATACCGTAT GTGCGTTTGA TCAGTTTTTC AGTGAAGGTT TCTTTTTTCA TGAGTTTGCT	3240
CCTTTTTTAA AAATCTTCCT CCCAAAAGAG ACTGTTGAGG TCAGTTTGGA GGCTGCGGGC	3300
GAGATTGAGA CAGAGTTCCA AGGTTGGATT GTACTTGTCG TTTTCAATCA TATTGATAGT	3360
CTGTCTCGAG ACACCGATAT CCTTGGCGAG TTCGAGCTGG GAAATACCCA ATTCTTGC	3420
AAATCTTTC ACACGATTCA TCTGTTCTCC TTTCTGATTT ATGTCGTATA TATTGACTA	3480
TATTATAGTC TTTTAAACAT AAAGTGTCAG GTATTTTTGA CATATTTTTT GAAGAAATAG	3540
TAGTCTCCTT GTCCTATTTG TCTGACAAGT GCAAGCTGGT CGGATTTGTG GTAAAAATAG	3600
TAAGATATGA CAAAAGAATT TCATCATGTA ACGGTCTTAC TCCACGAAAC GATTGATATG	3660
CTTGACGTAA AGCCTGATGG TATCTACGTT GATGCGACTT TGGGCGGAGC AGGACATAGC	3720
GAGTATTAT TAAGTAAATT AAGTGAAAAA GGCCATCTCT ATGCCTTTGA CCAGGATCAG	3780
AATGCCATTG ACAATGCGCA AAAACGCTTG GCACCTTACA TTGAGAAGGG AATGGTGACC	3840
TTTATCAAGG ACAACTTCCG TCATTTACAG GCATGTTTGC GCGAAGCTGG TGTTCAGGAA	3900
ATTGATGGAA TTTGTTATGA CTTGGGAGTG TCTAGTCCTC AATTAGACCA GCGTGAGCGT	3960
GGTTTTCTT ATAAAAAGGA TCGGCCACTG GACATGCGGA TGAATCAGGA TGCTAGCCTG	4020
ACAGCCTATG AAGTGGTGAA CAATTATGAC TATCATGACT TGGTTCGTAT TTTCTTCAAG	4080
TATGGAGAGG ACAAATCTC TAAACAGATT GCGCGTAAGA TTGAGCAAGC GCGTGAAGTG	4140
AAGCCGATTG AGACAACGAC TGAGTTAGCA GAGATTATCA AGTTGGTCAA ACCTGCCAAG	4200
GACTCAAGA AGAAGGGGCA TCCTGCTAAG CAGATTTTCC AGGCTATTTC AATTGAAGTC	4260
AATGATGAAC TGGGAGCGGC AGATGAGTCC ATCCAGCAGG CTATGGATAT GTTGGCTCTG	4320
GATGGTAGAA TTTCACTGAT TACCTTTCAT TCCTTAGAAG ACCGCTTGAC CAAGCAATTG	4380
TTCAAGGAAG CTTCAACAGT TGAAGTTCCA AAAGGCTTGC CTTTCATCCC AGATGATCTC	4440
AAGCCCAAGA TGGAATTGGT GTCCCGTAAG CCAATCTTGC CAAGTGCGGA AGAGTTAGAA	4500
GCCAATAACC GCTCGCACTC AGCCAAGTTG CGCGTGGTCA GAAAAATTCA CAAGTAAGAG	4560
GGAAAAAGAT GGCAGAAAAA ATGGAAAAA CAGGTCAAAT ACTACAGATG CAACTTAAAC	4620
GGTTTTCGCG TGTGGAAAAA GCTTTTACT TTTCCATTGC TGTAACCACT CTTATTGTAG	4680
CCATTAGTAT TATTTTTATG CAGACCAAGC TCTTGCAAGT GCAGAATGAT TTGACAAAAA	4740
TCAATGCGCA GATAGAGGAA AAGAAGACCG AATTGGACGA TGCCAAGCAA GAGGTCAATG	4800

223

AACTATTACG TGCAGAACGT TTGAAAGAAA TTGCCAATTC ACACGATTG CAATTAAACA 4860
ATGAAAATAT TAGAATAGCG GAGTAAGATA TGAAGTGGAC AAAAAGAGTA ATCCGTTATG 4920
CGACCAAAAA TCGGAAATCG CCGGCTGAAA ACAGACGCAG AGTTGGAAAA AGTCTGAGTT 4980
TATTATCTGT CTTTGTTTTT GCCATTTTTT TAGTCAATTT TCGCGTCATT ATTGGGACAG 5040
GCACTCGCTT TGGAACAGAT TTAGCGAAGG AAGCTAAGAA GGTTCATCAA ACCACCCGTA 5100
CAGTTCCTGC CAAACGTGGG ACTATTTATG ACCGAAATGG AGTCCCGATT GCTGAGGATG 5160
CAACCTCCTA TAATGTCTAT GCGGTCATTG ATGAGAACTA TAAGTCAGCA ACGGGTAAGA 5220
TTCTTTACGT AGAAAAACA CAATTAAACA AGGTTGCAGA GGTCTTTCAT AAGTATCTGG 5280
ACATGGAAGA ATCCTATGTA AGAGAGCAAC TCTCGCAACC TAATCTCAAG CAAGTTTCCT 5340
TTGGAGCAAA GGGAAATGGG ATTACCTATG CCAATATGAT GTCTATCAAA AAAGAATTGG 5400
AAGCTGCAGA GGTCAAGGGG ATTGATTTTA CAACCAGTCC CAATCGTAGT TACCCAAACG 5460
GACAATTTGC TTCTAGTTTT ATCGGTCTAG CTCAGCTCCA TGAAAATGAA GATGGAAGCA 5520
AGAGCTTGCT GGGAACCTCT GGAATGGAGA GTTCCTTGAA CAGTATTCTT GCAGGGACAG 5580
ACGGCATTAT TACCTATGAA AAGGATCGTC TGGGTAATAT TGTACCCGGA ACAGAACAAG 5640
TTTCCCAACG AACGATGGAC GGTAAGGATG TTTATACAAC CATTTCCAGC CCCCTCCAGT 5700
CCTTTATGGA AACCAGATG GATGCTTTTC AAGAGAAGGT AAAAGGAAAG TACATGACAG 5760
CGACTTTGGT CAGTGCTAAA ACAGGGGAAA TTCTGGCAAC AACGCAACGA CCGACCTTTG 5820
ATGCAGATAC AAAAGAAGGC ATTACAGAGG ACTTTGTTTG GCGTGATATC CTTTACCAAA 5880
GTAACATGA GCCAGGTTCC ACTATGAAAG TGATGATGTT GGCTGCTGCT ATTGATAATA 5940
ATACCTTTCC AGGAGGAGAA GTCTTTAATA GTAGTGAGTT AAAAATTGCA GATGCCACGA 6000
TTCGAGATTG GGACGTTAAT GAAGGATTGA CTGGTGGCAG AACGATGACT TTTTCTCAAG 6060
GTTTTGCACA CTCAAGTAAC GTTGGGATGA CCCTCCTTGA GCAAAAGATG GGAGATGCTA 6120
CCTGGCTTGA TTATCTTAAT CGTTTTAAAT TTGGAGTTCC GACCCGTTTC GGTTTGACGG 6180
ATGAGTATGC TGGTGAGCTT CCTGCGGATA ATATTGTCAA CATTGCGCAA AGCTCATTTG 6240
GACAAGGGAT TTCAGTGACC CAGACGCAA TGATTCGTGC CTTTACAGCT ATTGCTAATG 6300
ACGGTGTCAT GCTGGAGCCT AAATTTATTA GTGCCATTTA TGATCCAAAT GATCAAACCTG 6360
CTCGGAAATC TCAAAAAGAA ATTGTGGGAA ATCCTGTTTC TAAAGATGCA GCTAGTCTAA 6420
CTCGGACTAA CATGCTTTTG GTAGGGACGG ATCCGGTTTA TGGAAACCATG TATAACCACA 6480
GCACAGGCAA GCCAACTGTA ACTGTTCTG GGCAAAATGT AGCCCTCAAG TCTGGTACGG 6540

224

CTCAGATTGC TGACGAGAAA AATGGTGGTT ATCTAGTCGG GTTAACCGAC TATATTTTCT	6600
CGGCTGTATC GATGAGTCCG GCTGAAAATC CTGATTTTAT CTGTATGTG ACGGTCCAAC	6660
AACCTGAACA TTATTCAGGT ATTCAGTTGG GAGAATTGTC CAATCCTATC TTGGAGCGGG	6720
CTTCAGCTAT GAAAGACTCT CTCAATCTTC AAACAACAGC TAAGGCTTTA GAGCAAGTAA	6780
GTCAACAAAG TCCTTATCCT ATGCCTAGTG TCAAGGATAT TTCACCTGGT GATTTAGCAG	6840
AAGAATTGCG TCGCAATCTT GTACAACCCA TCGTTGTGGG AACAGGAACG AAGATTAAAA	6900
ACAGTTCTGC TGAAGAAGGG AAGAATCTTG CCCCGAACCA GCAAGTCCTT ATCTTATCTG	6960
ATAAAGCAGA GGAGGTCCA GATATGTATG GTTGACAAA GGAGACTGCT GAGACCTTG	7020
CTAAGTGGCT CAATATAGAA CTTGAATTTC AAGGTTGGG CTCTACTGTG CAGAAGCAAG	7080
ATGTTCTGTC TAACACAGCT ATCAAGGACA TAAAAAAT TACATTAACT TTAGGAGACT	7140
AATATGTTTA TTCCATCAG TGCTGAATT GTGACATTT TACTAACTT AGTAGAAAT	7200
CCGGCTTTA TCCAATTTA TAGAAAGCG CAAATTACAG GCCAGCAGAT GCATGAGGAT	7260
GTCAAACAGC ATCAGGCAAA AGCTGGGACT CCTACAATGG GAGGTTGGT TTCTTGATT	7320
ACTTCTGTT TGGTTGCTT CTTTTCGCC CTATTTAGTA GCCAATTCAG CAATAATGTG	7380
GGAATGATTT TGTTCATCTT GGTCTGTAT GGCTTGGTCG GATTTTGA TAGACTTCTC	7440
AAGGTCCTTC GTAAATCAA TGAGGGGCTT AATCCTAAGC AAAAATTAGC TCTTCAGCTT	7500
CTAGGTGGAG TTATCTTCTA TCTTTCTAT GAGCGCGGTG GCGATATCCT GTCTGTCTT	7560
GGTTATCCAG TTCATTTGGG ATTTTCTAT ATTTTCTCG CTCTTTCTG GCTAGTCGGT	7620
TTTTCAAACG CAGTAACTT GACAGACGGT GTGACGGT TAGCTAGTAT TTCCGTTGTG	7680
ATTAGTTTGT CTGCCTATGG AGTTATTGCC TATGTCAAG GTCAGATGGA TATTCTTCTA	7740
GTGATTCTTG CCATGATTGG TGGTTGCTC GGTTCCTCA TCTTTAACCA TAAGCCTGCC	7800
AAGGTCCTTA TGGGTGATGT GGAAGTTTG GCCCTAGGTG GGATGCTGGC AGCTATCTCT	7860
ATGGCTCTCC ACCAAGAATG GACTCTCTG ATTATCGGAA TTGTGTATGT TTTTGAAACA	7920
ACTTCTGTTA TGATGCAAGT CAGTTATTTT AACTGACAG GTGGTAAACG TATTTCCGT	7980
ATGACGCCTG TACATACCA TTTTGAGCTT GGGGATTGT CTGGTAAAGG AAATCCTTGG	8040
AGCGACTGGA AGGTTGACTT CTTCTTTTGG GGAGTGGAC TTCTAGCAAG TCTCCTGACC	8100
CTAGCAATTT TATATTTGAT GTAAGAATGG CACCCTGATG TTTCAGGG	8148

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9909 base pairs
 (B) TYPE: nucleic acid

225

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TACTCCACCC TTAATATCCG TTCCTGTAAA TACTTTACCG CTTTAAAGTT CATAGAATTG	60
AACCTTTAAA TGCTTGTCTT CAAGCATCTT TTCCATCCAA TTTTtaggag TTGACCAGC	120
TTTAAATAAA AACCTTGCTG GGGTGATTAG TATAGATTTA TCTGCGATTT TATAAGCTTC	180
ATCAATAAAA TAGTGATATA TCGGCTCATC TCTGGCTTCT CCTGTTTCCT GATACGGAGG	240
ATTTCTATC ACGACATCAA ATTTCAATTC ACTTTCCTCG CTAGATAGGC GCTCAAACC	300
TATCATTCTA TTCTTTTCC AGTCTTTGAT ATGGGTTTGA GATTCTTCTA CTCTTGGAC	360
TTCTAGCTCA TCCGCAAACA AACTCAATTG TTGAGATTGC TTTTGTtag CTGAATAAGG	420
ACTACTTTTT TTCAATCCAT CCATCTGAAA GACATTGTAA GAGATAATAG TCGCAATTC	480
TTCTTTTGC TCTAATGTTG GTTGATTTC AGTCTTAGCT AGATAATAGT CCTCAAAGT	540
TGCCAAAAGA TTCTCAGCG CCAAAGGAG AGAATCTCCT TGATACTCAT AACCATACGA	600
AGCATGATAA GCATCTTTTA CAAGTTTATA AAATGTGACT TCATCTGAAA CCTCAGACT	660
AATCCGTTGC AGTTTCTAT CAACAAAACC AACTCGCTCA GATAATGGAA TTCTCTCACC	720
AGTTACGGTA TCATATCTCG TTACCATATA AGGTGCTTCA CCACAAGTTA CCTCTAACCA	780
TCGTAAGTCC ACATACTCCT CAAGACTTAA CGAGCCTAAT TCGATTCTA CATATCCATT	840
TTGCTTTGCG ACCAACCACG TTGGTGTAAG CACTCTGCC CTTATTTTGG TCCGATCTTT	900
TTGTTTCATAT TTGGATTTTT CAGATCTGGG CTGAATCAAG TTGGCAAAGT TTCCAGTAAC	960
CTTACTTGGA TTGATGCGAT CACTTGAGC AAATCCCTTT CTAACAATT CATAAGAATG	1020
CGTANGCCAA ACAATTGATT TCTTTGTCGT TCGATCTTTT AAAAGAATTT TTAATAAGTC	1080
AGCCGATTCT TTAGCCAAAC TTTCTTCACT AATATCTATT GTCATCAGCA ACCTCTCTTA	1140
TATTGTAAGC CCTATTATAT CATATTTTAA AGAATGAAAA TTTACTTGAA AAAAGTAATT	1200
CAATAAATAT CTCTCCGATG ACCAACTTCT AGAGTAGCAA CGACTAATTC ATCATCTACA	1260
ATTGTACGA TAACTCGATA ATTACCAATT CTATAGCGCC ATTGACCAAC GCGATTACCA	1320
ACCAAAGCCT TTCCGTGTCG TCTTGGGTCT TCCAAAACAT TGGTTTGTA ATAGTTTGTA	1380
ATTAGCTTCT GCGTATAACG GTCCAATTTT TTCAATTGCT TGATAAAACG TCTTGTGGA	1440
ACTAATTTAT ACAAATTATT CATCCTTCAA GCCTAAATCA TGCATCATTT CTCCCAAGT	1500
AATGGGTTCA ACTCCTTTT CCAAGTCTTC TAAATACTCT TGATAGGCTA AATCTGCCAC	1560

226

ACGAGCATCG TATTCATCTT CTAGGGCTTC AAGAGTTTGT GTGCGAATAA GTTCCGAAAG	1620
GGAAACTCCT TCAAACTTAG CCATTGCTTT CATAAATGTT TTATCAGCTT CAGAAACTTT	1680
TAATGTAATA GTAGTCATCT TTTGTGCTCC CTTTTTTAAT GGTAAACACCA TTGTATTACT	1740
TTTTAGGTGT TCAGTCAATA TAAAAAGAAC ACCTTCTCAG CGTTCTTTCT ATATCTCTGT	1800
CAATGGTGTG GCGGTATCTG GTGAGGTATC ATAAACCTTA AAGTCTACTC CGACTCCAG	1860
ATCAGCTTGA GCCAGCTGAT TGACCATGGT CATATGAGCC AGTTCCTTGA TATTGTTTTC	1920
CTTAGATAAA TGCCCAAGGT AAATCTTCTT AGTACGATTT CCTAGCGTCC GAATCATAGC	1980
TTCAGCACCG TCCTCGTTAG AAAGGTGACC AAGGTCAGAT AGGATTCGTT GTTTGAGTCG	2040
CCAAGCGTAA GAACCTGATC GCAAAATCTC TACATCATGG TTGGCCTCGA TAAGATAACC	2100
ATCCCGCATTT TCGACAATGC CCGCCATACG GTCAGTACA TAACCTGTAT CTGTCAAGAG	2160
GACAAAATC TTATCATCCT TCATAAAGCG ATAGAACTGC GGTGCGACTG CATCATGGCT	2220
TACACCAAAA CTCTCGATGT CGATATCTCC AAAGGTTTTG GTTTTACCCA TTTCAAAAAT	2280
ATGCTTTTGC GAAGAATCCA CCTTGCCAAG ATATTTACTA TTTTCCATAG CTTGCCAGGT	2340
CTTTTCATG GCATAAGAT CCATACCATA CTTGCGAGCC AAAACGCCTA CTCCATGGAT	2400
ATGATCTGAA TGCTCATGGG TAATCAAGAT GGCATCCAGG TCTTCTGGCT TACGGTTAAT	2460
TTTCTAGTGC AGACTGGTAA TTTTCTTGCC AGACAAGCCT GCATCTACTA AAAGCTTCTT	2520
TTTTGAGGTT TCCAGATAAA AAGAATTTC ACTGGAACCC GACGCTAAAA TACTGTATTT	2580
AAAGCCTATT TCACTCATTC TAGTCTTCTA CTTTCATCCTC CCATACTTCT TCTTTCAGT	2640
CATCCTTATC ATAAGGGAGT ACAATGGTAA AGGTGAACC CTTGCCGTAT TCACTCTTGG	2700
CCCAAATAAA GCCCTTATGT TGTTTGATAA TTTCTTTAGC GATAGACAGT CCTAGACCTG	2760
TACCACCTTG TGCACGACTT CTAGCAGGAT CCACACGATA GAAACGGTCA AAGATACGTG	2820
GTAAATCCTG CTTAGGAATC CCCAAACCGT GGTCAGAAAT GGATAAAATC ATCTGGTCTT	2880
CAGTTGTCTT CATCTGACA GTGATTTTAC CCCCATCTGG CGAATACTTA ATAGCATTAT	2940
TTAAAATATT GTCGACAACC TGCATCATCT TATCTGTATC AATTTCCATC CAGATAGAAT	3000
TGATGGGATA ATCTCTCACC AACTCATATT TTTTCTCCTT TTCCTGTCCT TTCATCTGT	3060
CAAAACGATT GAGGATAAAG GTAATAAAG CAGTGAAGTT AATCAGTTCC ACATCTAGGT	3120
GACTGGTAGC ATATCAATA CGTGAAAGAT GGAGGAGATC CGTCACCATG CGCATCATAC	3180
GGTTGGTCTC ATCAAGAGAA ACCTTGATAA AGTCTGGTGC TACAGTTTCA CACAAAGCCC	3240
CCTCATCCAA GGCTTCAAGA TAGGATTTTA CGCTAGTCAG AGGAGTCCGT AACTCATGGC	3300
TAACATTGGA AACAAAGAGT CTTCTGTCGC GTTCTTCTT CTCCTGCTCC GTCGTATCAT	3360

227

GCAAAACAGC CACCAAACCT GAAATAAAGC CAGACTCTCG ACGTATCAAG GCAAAGCGAA	3420
CTCGAAGGTT CAAATATTCG CCATTGATAT CTTGGGAATC TAGCAACAAT TCTGGACTTT	3480
GGGTAATCAA ATCACGCAAT TCATAGTTTT CTTCTATCTT GAGCAATTCC AAAATGCTTC	3540
TATTCAGAAC ATCTTCCTTA ACCAACCCCA GTTGCTTCTT GGCTGTATCG TTAATCATGA	3600
TAATCTGACC CCGACGGTTA GTCGAAGAA CCCCATCTGT CATATAAAAC AGAATACTAT	3660
TTAGCCTCTT ACTCTCTTGT TCTAGATTTT CCTGAGTGAG ACGAATAACC TCCGACAAGT	3720
CATTCAAATT ATTGGTAATA TTGGTGATTT CAGACCCACC TTGCATATCA AGAACCTTGG	3780
AATAATCTCC TGCAATCAA TCTTTAACCT TTTGATTGAC TTGCTTCAAC TGAATATTAT	3840
CACGTCTATT TTCCAGTAAT AAGAGGGTCA CAACAAGGAT GAAACCTAAC AAAATCAGGA	3900
TAAAGATAAA ATCTCTGGTA AAAATGGTTT GTTTCAGTAA ATCAAGCATT ATTTCTCATG	3960
TAATACCCTA CACCACGGCG CGTCAAGATA TACTCTGGTC GGCTGGGCGT ATCTTCAATC	4020
TTCTCACGCA GACGTCGTAC AGTCACATCA ACTGTACGGA CATCACCAA ATAGTCATAA	4080
CCCCAGACAG TCTCAAGCAA GTGTCGCGC GTGATGACTT GACCTGTATG CGATGCTAAA	4140
TGATACAAAA GCTCAAATTC ACGATGGGTT AAGTCTAGTT CTTCCGCCATA TTTTITAGCC	4200
ACGTAGGCGT CTGGAACAAT TTCTAAATCC CCAATTGGA TAGGTTGAGG TTTACTATCT	4260
GCTTCCTGAC CATCTACTGG CATAGGTTGA GAACGACGCA GAAGAGCTTT AACACGCGCC	4320
TGCAACTCAC GATGGAGAA GGGTTTGTG ACATAGTCAT CTGCCCAAG TTCCAAACCG	4380
ATAACCTTAT CAAATTCACT ATCTTTGGCT GAAAGCATAA GAATGGGCAC ACTGCTTGTC	4440
TTACGAATGG TCTTAGCAAC TTCTAAACCA TCAATTTCTG GAAGCATCAA ATCCAGAATA	4500
ATAATATCTG GTTGCTCTGC TTCAAATTGC TCTAGCGCTT CACGACCATT AAAAGCAGTT	4560
ACAACTTCGT AACCTTCCTT GGTATATTA AACTTGATAA TATCCGAGAT TGGTTTCTCA	4620
TCATCTACAA TTAGTATTTT TTTCATATGT TCACCTTTTT CTCTACTATT ATACCAAAAA	4680
AATAGTCAGA AGACACAATA GCTAGTCTTG GCTACTGTCT AAGTTGGCTT GTGCATAAAC	4740
CTGCCAGATT TTTTGTGGG GTTGGCAAG TGGGTAATTC TTGAATTCTT CTGGTGAAAG	4800
CCAGCGAACT TCCCTATCTG AAAAATCATG GAAGTCACTC ACCTGACCTG CTACAATCTG	4860
TACATGCCAT TTTCGATGAC TAAAAACATG CTGGACTGTA TCAAAACAAA CATCAAGCCA	4920
ATCAACATCT AGGTCATAGT CCTGCTGGAA ACTCTCTTCT GGACTGGGAC CAAAGTTCAC	4980
ACTTCTTCC GCAACCTGAT GAAAGAGGTC AAAGTGTCT TCTTGGGAAA AGTTATCAAC	5040
TTCTATAAAG GGGAAATGCC AAAACCTGC CAAGAGCTTT TCGCTTTCAT TTTTTTCAAG	5100

228

TAAAAATGCT CCTTGAGAAT TTTTCACAAC TAAGGCTTTA AGATAAATAG GAACCGGCTT	5160
TTTCTTAGGA GATTTAATTG GATAACGGTC CATGGTTCCA TTCTGATATG CCGCACTAAA	5220
GTCTTGAAGT GGGCTTTCTT CAGGTCTGGG ATTTACAGGA GACTCAATAT CAGACCCTAA	5280
GTCCATCAAG GCTTGATTAA AATCACCCGG ACGATCCGGA TTAATCAAGA TCTCCATCAT	5340
TGCCTGAAAA ATTITTCGAT TACTTGGAAT CCCAATATCG TGGTTGACTT CAAACAGACC	5400
CGCCAAGACC CGCATGACAT TACCATCTAC AGCTGGCTCA GGCAAGTTAA AAGCAATACT	5460
GGAAATGGCT CTGCTGTGT AAGGTCCAAT CCCTTTCAAG CTGGAAATTC CTTCATAGGT	5520
ATTTGGAAT TGGCCACCAA AGTCAGTCAT AATCTGCTGG GCTGCAGCCT GCATATTGCG	5580
AACTCGAGAA TAATAGCCCA AGCCCTCCCA AGCTTTCAGT AAATCTCCT CAGGCGCAGT	5640
TGCCAGACTT TCGACAGTTG GAAACCAGTC CAAAATCTT TCGTAGTAAG GGATAACTGT	5700
ATCCACCCTG GTCTGCTGAA GCATGATTC AGATACCCAG ATGTGATAAG GATTTTACT	5760
TCTCCTCAA GGCAAATCTC TTTGTGTTTC ATCATAACAA GCGAGAAGTT TCTCACGGAA	5820
AGAAATGACT TTCTCCTCCG GCCACATGAC GATACCGTAT TCTTTCAAAT CTAACATATC	5880
TCTAGTATAA CACAGAAGGT TTCACCTGTC TTTGTATCTG ATTTATAATA TTTTCAATAG	5940
ATAGTATATA ACTTTTCTAT CTACTTATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAG	6000
CTAGCCGAG GTTGCTCAAA AACTGTGTTT GAGGTGTGG ATAGAACTGA CAGAGTCAGT	6060
ATCATATACT ACGGCAAGGT GAAGCTGACG TAGTTTGAAG AGATTTTCGA AGAGTATAAA	6120
TCTTATTGAT GAACTGCTTG CAGTCTGAGA AAAAATGAGC TTGGATATTA TTTCCAAACT	6180
CACTTAAAGT CAATTTCAAT CCACTAGAAC AAGCCTAGTA CAGTTCCATC GCTTTCAACA	6240
TCCATGTTGA GAGCTGCTGG ACGTTTGGGA AGACCTGGCA TGGTCATAAC ATCACCAGTT	6300
AAGGCAACGA TGAAGCCTGC ACCTAATTTT GGTACCAATT CACGAATGGT AATTTCAAAG	6360
TTTTCTGGTG CTCCAAGCGC ATTTGGATTG TCTGAGAAAC TGTATTGAGT TTTAGCCATA	6420
CAGATTGGCA ATTTGTCCCA ACCGTTTGA ACGATTGAG CAATTTGTGT TTGAGCTTTC	6480
TTCTCAAAGT TCACTTTGCT ACCACGATAG ATTTCACTGA CAATTTTTC AATCTTTTCT	6540
TGGACAGAAA GGTCAATTATC ATACAAACGT TTATAGTTAG CTGGATTTC AGCAATTGTC	6600
TTAACAAC TGCTACTCCA CCTTCTGCTC CATCAGCCCA GACACTAGCC	6660
AATTCAACTG GTACATCGAT TGAGGCACAG AGTTCTTTTA AGGCTGCAAT TTCAGCTTCT	6720
GTATCAGATA CAAATTCGTT AATAGCTACA ACTGCTGGAA TACCGAACTT ACGGATATTT	6780
TCAACGTGGC GTTTCAAGTT AGCAAAACCT GCACGAACTG CCTCTACATT TTCTTCAGTC	6840
AGAGCGTCTT TAGCCACACC ACCATTATC TTAAGGGCAC GAAGGTTGC GACAATAACA	6900

ACTGCATCTG GAGATGTTGG CAAGTTTGGT GTCTTGATAT CAAGGAATTT CTCAGCACCA	6960
AGGTCCGCAC CAAAACCAGC TTCAGTAACA GTGTAATCAG CCAAGTGAAG GGCTGTTGTC	7020
GTGCCCCAAA CAGAGTTACA GCCATGAGCG ATATTGGCAA ATGGACCACC GTGTACAAAG	7080
GCAGGTGTAC CGTAAATTGT CTGAACCAAG TTTGGCTTAA TAGCATCCTT CAAAATCAAA	7140
GCCAAGGCAC CCTCAACCTG CAAATCACCT ACAGAAACAG GCGTACGGTC ATAGCGATAA	7200
CCAATAACGA TATTCGCCAA ACGACGTTTC AAGTCCTCGA TGTCCGTTGC CAAGCAAAGA	7260
ATTGCCATGA TTTCTGAAGC AACTGTAATA TCAAAACCAT CCTCACGTGG AATACCGTTT	7320
AGAGGACCAC CAAGACCAAC AGTCACATGG CGGAGCGTAC GGTCCGTTCAA GTCCACAACG	7380
CGTTTCCAGA GGATACGACG TTGATCAATT CCCAGCTCAT TCCCTTGGTG CAAGTGGTTG	7440
TCAATCAAGG CAGAAAGGGC ATTGTTGGCA GTTGTAAATAG CATGCATATC TCCAGTAAAG	7500
TGGAGGTTGA TGTCTTCCAT TGGCAGAACT TGTGCATACC CACCACCAGC AGCACCACCC	7560
TTGATCCCCA TGA CTGGACC AAGAGACGGT TCGCGGATAG CAATCATGGT TTTCTTGCCA	7620
ATCTTGTTC AAGGCATCCGC AAGACCAATG GTAAGCGTCG ACTTTCCTTC ACCTGCAGGT	7680
GTGGGGTTGA TGGCAGTAAC CAAGATCAAT TTACCGACTG GATTGCTCTC AACTGCACGA	7740
ATTTTATCAA AGCTGAGTTT AGCCTTGATC TTCCGTACA ACTCCAAATC GTCATAAGAA	7800
ATACCAAGTT TCTCTACAAC ATCAACAATT GGCTTCAACT CAATACTCTG TGCGATTTC A	7860
ATATCTGTTT TCATTCAAAA TTCCTCTAAC CTCTTATATG ATAATTCAAT ATATCACAAA	7920
ACAAGATTTT TAACATCCTA AACTCTCTA AACGTCGTA AATATCTCTG TTTTAAAGAC	7980
TTT TAGAGTC CTTTCTTAAA TTTTATATGG CTTTATAGTT TGAAACTATA ATAAATCTTC	8040
GTTTTTACCA AAAATTTATC ACTTTCATTT TACTTACCGC TTATTTTGT GTACAATAGT	8100
GCTATGAAAA TTTTAGTTAC ATCGGGCGGT ACCAGTGAAG CTATCGATAG CGTCCGCTCT	8160
ATCACTAACC ATTCTACAGG TCACTTGGGG AAAATTATCA CAGAGACTTT GCTTCTGCA	8220
GGGTATGAAG TTTGTTAAT TACGACAAAA CGAGCTCTGA AGCCAGAGCC TCATCCTAAC	8280
CTAAGTATTC GAGAAATTAC CAATACCAAG GACCTTCTAA TAGAAATGCA AGAACGTGTT	8340
CAGGATTATC AGGTCTTGAT CCACTCAATG GCTGTTCTG ACTACACTCC TGTATTATG	8400
ACAGGGCTTG AGGAAGTTCA GGCTAGCTCC AATCTAAAAG AATTTTAAAG CAAGCAAAAT	8460
CATCAGGCCA AGATTCTTC AACTGATGAG GTTCAGGTTT TGTTCCTTAA AAAGACACCC	8520
AAAATCATAT CCCTAGTCAA GGAATGGAAT CCTACTATTC ATCTGATTGG TTTCAAACCTG	8580
CTGCTTGATG TTACCGAAGA TCATCTGGTT GACATTGCAC GAAAAAGTCT TATCAAGAAT	8640

230

CAAGCAGATT TAATCATCGC GAATGACCTG ACTCAAATTT CAGCAGATCA GCACCGAGCT	8700
ATATTTGTTG AGAAAAATCA GCTTCAAACA GTCCAGACTA AAGAAGAAAT TGCAGAACTC	8760
CTCCTTGAAA AAATTCAAGC CTATCATTCT TAGAAAGGAA AACTATGGCA AACATTCTCT	8820
TGGCTGTAAC GGGTTCAATC GCCTCTTATA AGTCGGCAGA TTTAGTCAGT TCTCTAAAAA	8880
AACAAGGCCA TCAAGTCACT GTCTTAATGA CTCAGGCTGC TACAGAGTTT ATCCAACCTT	8940
TGACACTACA GGTACTCTCA CAGAATCCTG TCCACTTGGA TGTCATGAAG GAACCCATATC	9000
CTGATCAGGT CAATCATATC GAACTTGGA AAAAAGCAGA TTTATTATATC GTGGTACCTG	9060
CAACTGCTAA CACTATTGCA AAAC TAGCTC ACGGATTTGC GGACAACATG GTAACCAAGTA	9120
CAGCTCTAGC CCTACCAAGT CATATCCCA AACTAATAGC TCCTGCTATG AATACAAAAA	9180
TGTATGACCA TCCAGTAACT CAGAATAATC TGAAAACATT AGAAACTACG GCTATCAGCT	9240
GATTGCTCCT AAGGAATCCC TACTAGCTTG TGGAGACCAC GGACGAGGAG CTTTAGCTGA	9300
CCTCACAATT ATTTTAGAAA GAATAAAGGA AACTATCGAT GAAAAAACGC TCTAATATTG	9360
CACCCATGTC TATCTTTTTC GCTACCATGC TCGTGATACA CTTTCTGAGC TCACCTTATCT	9420
TTAACCTTTT TCCATTTCCA ATCAAACCGA CCATTGTTCA TATTCCTGTC ATTATTGCCA	9480
GCATTATTTA TGGTCCACGA GTTGGGGTTA CACTTGGATT TTTGATGGGA TTAATTAGCT	9540
TGACGGTTAA CACGATTACG ATTCTACCGA CAAGCTACCT CTTCTCTCCC TTCGTACCAA	9600
ACGGAACAT CTACTCAGCT ATCATTGCCA TCGTCCCACG TATTTTGATT GGTTTAACTC	9660
CTTACTTAGT CTATAAATG ATGAAAAACA AGACTGGTCT GATTTTAGCT GGAGCCCTTG	9720
GTTCCTTGAC AAATACTATC TTTGTCCTTG GAGGAATCTT CTTCTATTT GGAAATGTTT	9780
ATAATGGAAA TATCCAACTT CTTCTGGCAA CCGTTATCTC AACAAATCA ATTGCTGAAT	9840
TGGTCATTTT TGCAATTCTA ACCCTAGCCA TTGTTCCACG ACTACAAACC TTGAAAAAAT	9900
AAAAACAGG	9909

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1126 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TAATTTTCAT ATAATAGTAA AATAGAATGT GTGATTCAT AATCACCTCA AATAGAAAGG	60
AAATTCTATG TCAAATCTAT CTGTTAATGC AATTCGTTTT CTAGGTATTG ACGCCATTAA	120

231

TAAAGCCAAC TCAGGTCATC CAGGTGTGGT TATGGGAGCG GCTCCGATGG CTTACAGCCT	180
CTTTACAAAA CAACTTCATA TCAATCCAGC TCAACCAAAC TGGATTAACC GCGACCGCTT	240
TATTCTTTCA GCAGGTCATG GTTCAATGCT CCTTTATGCT CTTCTTCACC TTTCTGGTTT	300
TGAAGATGTC AGCATGGATG AGATTAAGAG TTTCCGTCOA TGGGGTTCAA AAACACCAGG	360
TCACCCAGAA TTTGGTCATA CGGCAGGGAT TGATGCTACG ACAGGTCCTC TAGGGCAAGG	420
GATTTCAACT GCTACTGGTT TTGCCCAAGC AGAAGCTTTC TTGGCAGCCA AATATAACCG	480
TGAAGGTTAC AATATCTTG ACCACTATAC TTACGTTATC TGTGGAGACG GAGACTTGAT	540
GGAAGGTGTC TCAAGCGAGG CAGCTTCATA CGCAGGCTTG CAAAACTTG ATAAGTTGGT	600
TGTTCTTTAT GATTCAAATG ATATCAACTT GGATGGTGAG ACAAAGGATT CCTTTACAGA	660
AAGTGTTCGT GACCGTTACA ATGCCTACGG TTGGCATACT GCCTTGGTTG AAAATGGAAC	720
AGACTTGGAA GCCATCCATG CTGCTATCGA AACAGCAAAA GCTTCAGGCA AGCCATCTTT	780
GATTGAAGTG AAGACGGTTA TTGGATACGG TTCTCCAAAC AAACAAGGAA CTAATGCTGT	840
ACACGGCGCC CCTCTTGAG CAGATGAAAC TGCACTCACT CGTCAAGCCC TCGGTTGGGA	900
CTACGAACCA TTTGAAATTC CAGAACAAGT ATATGCTGAT TTCAAAGAAC ATGTTGCAGA	960
CCGTGGCGCA TCAGCTTATC AAGCTTGGAC TAAATTAGTT GCAGATTATA AAGAAGCTCA	1020
TCCAGAACTG GCTGCAGAAG TAGAAGCCAT CATCGACGGA CGTGATCCAG TCGAAGTGAC	1080
TCCAGCAGAC TTCCAGCTT TAGAAAATGG TTTTtCTCAA GCAACT	1126

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2520 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CCGGCAACAA AAAAGAAAAA ATCAACAGTT AAAAAAATC TAGTCATCGT GGAGTCGCCT	60
GCTAAGCCAA GACGATTGAA AAATATCTAG GCAGAACTA CAAGGTTTGA GCCAGTGTCTG	120
GGCATATCCG TGATTGAAG AAATCCAGTA TGTCCGTCGA TATTGAAAAA AATTATGAAC	180
CGCAATATAT TAATATCCGA GGAAAGGCC CTCTTATCAA TGACTTGAAA AAAGAAGCTA	240
AAAAGCTAA TAAAGTTTTT CTCGCGAGTG ACCCGGACCG TGAAGGAGAA GCGATTTCTT	300
GGCATTTGGC CCATATTCTC AACTTGGATG AAAATGATGC CAACCGTGTG GTCTTCAATG	360

232

AAATCACCAA GGATGCAGTC AAAAATGCTT TTAAAGAACC TCGTAAGATC GATATGGACT 420
TGGTCGATGC CCAACAAGCT CGTCGGATCT TGGATCGCTT GGTAGGGTAT TCGATTTCGC 480
CTATTTTGTG GAAGAAGGTC AAGAAGGGCT TGTACGAGG TCGCGTTCAG TCCATTGCC 540
TTAAACTCAT CATTGACCGT GAAAATGAAA TCAATGCCTT CCAGCCAGAA GAATACTGGA 600
CAGTTGATGC TGTCTTTAAA AAGGGAACCA AACAAATTCA TGCTTCCTTC TATGGAGTAG 660
ATGGTAAAAA GATGAACTG ACCAGCAATA ACGAAGTCAA GGAAGTCTTG TCTCGTCTGA 720
CGAGTAAAGA CTTTTCAGTA GATCAGGTGG ATAAGAAAGA GCGCAAGCGC AATGCTCCTT 780
TACCCTATAC CACTTCATCT ATGCAGATGG ATGCTGCCAA TAAAATCAAT TTCCGTA CT 840
GAAAAACCAT GATGGTTGCC CAACAGCTCT ATGAAGGAAT TAATATCGGT TCTGGTGTT 900
AAGGTTTGAT TACCTATATG CGTACCGATT CGACTCGTAT CAGTCTGTGA GCGCAAAATG 960
AGGCGGCAAG CTTCAATTACG GATCGTTTGG GTAGCAAGTA TTCTAAGCAC GGTAGCAAGG 1020
TCAAAAACGC ATCAGGTGCT CAGGATGCC ATGAGGCTAT TCGTCCGTCA AGTGTCTTTA 1080
ATACACCAGA AAGCATCGCT AAGTATCTGG ACAAGGATCA GCTTAAGCTA TATACCCTTA 1140
TCTGGAATCG TTTTGTGGCT AGCCAGATGA CAGCGGCCGT TTTTGATACC ATGGCTGTTA 1200
AATTGTCTCA AAAAGGGGTT CAATTGCTG CCAATGGTAG TCAGGTTAAG TTTGATGGTT 1260
ATCTTGCCAT TTATAATGAT TCTGACAAGA ATAAGATGTT ACCGGACATG GTTGTGGAG 1320
ATGTGGTCAA ACAGGTCAAT AGCAAACCAG AGCAACATTT CACCCAACCG CCTGCCCGTT 1380
ATTCTGAAGC AACACTGATT AAAACCTTAG AGGAAAATGG GGTGGGACGT CCATCAACCT 1440
ACGCGCCAAC CATTGAAACC ATTCAGAAAC GTTATTATGT TCGCCTGGCA GCCAAACGTT 1500
TTGAACCGAC AGAGTTGGGA GAAATTGTCA ATAAGCTCAT CGTTGAATAT TTCCCAGATA 1560
TCGTAAACGT GACCTTCACA GCTGAAATGG AAGGTAAACT GGATGATGTC GAAGTTGGAA 1620
AAGAGCAGTG GCGACGGGTC ATTGATGCCT TTTACAAACC ATTCTCTAAA GAAGTTGCCA 1680
AGGCTGAAGA AGAAATGGAA AAAATCCAGA TTAAGGATGA ACCAGCTGGA TTTGACTGTG 1740
AAGTGTGTGG CAGTCCAATG GTCATTAAAC TTGGTCGTTT TGGTAAATTC TACGCTTGTA 1800
GCAATTTCCC AGATTGCCGT CATACCCAAG CAATCGTGA AGAGATTGGT GTTGAGTGTC 1860
CAAGCTGTCA TCAGGGACAA ATTATTGAGC GAAAAACCA GCGTAATCGC CTATTCTATG 1920
GTTGCAATCG CTATCCAGAA TGTGAATTTA CCTCTTGGGA CAAGCCTGTT GGTCTGACT 1980
GTCCAAAATG TGGCAACTTC CTCATGGAGA AAAAAGTCCG TGGTGGTGGC AAGCAGGTTG 2040
TTTGTAGCAA AGGCGACTAC GAGGAAGAAA AGATGGCTCT TTGTCAACTG TAGTGGGTTG 2100
AAGTCAGCTA AGCTCGAGAA AGGACAAATT TTGTCCTTTC TTTTTTGATA TTCAGAGCGA 2160

233

TAAAAATCCG TTTTGAAG TTTTCAAAGT TCCGAAAACC AAAGGCATTG CGCTTGATAA	2220
GTTTGATGAG ATTATTGGTC GCTTCCAATT TGGCGTTAGA ATAGTGTAGT TGAAGGGCGT	2280
TGACGATTTT CTCTTTGTCC TTTAGAAAGG TTTTAAAGAC AGTCTGAAAA AGAGGATGAA	2340
CCTGCTTTAG ATTGTCTCA ATGAGTCCGA AAAATTCTC CGGTCCTTA TTCTGAAAGT	2400
GAAACAGCAA GAGTTGATAG AGCTGATAGT GATGTTTCAA GTCTTGTGAA TAGCTCAAAA	2460
GCTTGTTTAA AATCTCTTA TTGGTTAAAT GCATACGAAA AGTAGGGCGA TAAAAATGTT	2520

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10993 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TTTCTCGAT AATAACTTCC ACCTTATTAT TTGGGATACC TCCTCTTCT TCACCACCAC	60
GTTCATAGTA GTCATCGCGA TAGAGAAAAG CTACGATATC AGCGTCCTGC TCAATAGACC	120
CAGATTACAG AATATCAGAC AAGACCGGTC TCTGTCTCTG ACGTTGTTCT ACACCACGAG	180
AAAGCTGACT CAGAGCGATT ACTGGAACCT TCAATTCCTT GGCTAGTATT TTCAACTGAC	240
GAGAAATTC AGAAACTTCT TGTGACGAT TTTCTCGACC AGTCCCGTG ATAAGTTGCA	300
AATAGTCTAT CAAAATCAAA CCAAGATTTC CAGTTTCTTG AGCCAATTTA CGAGAACGAG	360
AACGAATCTC TGTAAATCCGA ATACCTGGCG TATCATCGAT ATAGATACTG GCGTTAGcTA	420
GATTACCCTG AGCAATAGTA TATTTTGGC ACTCCTCATC TGTC AATTGC CCTGTACGGA	480
TAGAAATGTA CTCCACTAAG CCTTCTGCAG CTAACATACG ATCTACCAAG CTTCCGCAC	540
CCATTTGAG TGAAAAAATA GCAACCGTTT TGTCCAACCT AGTCCCAATG TTCTGAGCGA	600
TATTCAAGGC AAATGCTGTC TTACCAACTG CTGGACGAGC TGCTAAGATA ATCAACTCCT	660
CCTCATGAAG TCCTGTTGTC ATATGATCCA AATCAGATA ACCTGTCGCA ATACCTGTAA	720
TATCGGTCGT TTGTTGCGAG CGAGCTTCCA GATTTCCAA GTTGAGATT AACACATCTC	780
GAATGTTCTT AAACCCGCTT CGATTGTCAT TTCTACTGAC ATCAATCAAC CCTTTTCTG	840
CCTGAGCAAT AATTCATCA GCTGGTTGTG ACGCTTCGTA AGCTTGTTG ACAGACTCTG	900
TCAACTTGGC AATTAAACGA CGTAGCATTG CTTTTCTGC AACAATCTTA GCATAATACT	960
CCGCATTAGC AGAAGTGGC ACAGAATTAA CAATCTCAAC CAAGTAAGAC AAGCCACCAA	1020

234

TATTCTGTAA ATCACCTGA TTATCAAGGA TAGTACGAAC CGTTGTTGCA TCTATGGCAT	1080
CACCACGATC GGATAAATCG ACCATGGCTT GGAAAATCAA ACGATGGGCA TACTTAAAAA	1140
AGTCCCGAGA CTCAATGTAT TCTCGCACAA AAACAAGTTT ACTCTCATCA ATAAAGATAG	1200
CCCCTAAAC GGATTGCTCA GCTAAGATAT CTTGAGGTTG TACTCGTAAC TCTTCTACTT	1260
CTGCCATCAG ACITCCCTTC CTTTACAAT CTTGTCAAGA AGGTGTAAAC TTATCCTTCT	1320
TTCACACGAA GATTGATTAC ACTTGTGATA TCTTGATAGA TTTTCACTGG CACATCAATC	1380
AAACCAACCG CTCGAATCGG AGCTGTACT TGAATATGAC GTTATCAAT CTTAATTTCCA	1440
AATTGCTTTT GCAATTCTTC TGCAATCTTC TTATTGGTAA TAGAACCAAA GGTACGACCA	1500
TCTGGACCAA CTTTTCACAA AAATTCTACA ACAGTTTCTT CTGCTTCAAG TTGTGCTTTA	1560
ATTGCTTTTC CTTCTGCAAT CATCTCAGCG TGAGCTTTT CTTCCGATTT TTGTTTACCA	1620
CGAAGTTTAC CTACAGCTTG AGCAGTCGCT TCTTGGCTA GATTCTTTT GATAAGAAAG	1680
TTTTGCGCAT ACCCTGTTGG TACTTCCTTA ATTTCGCCTT TTTTACCTTT TCCTTTAACA	1740
TCTGCTAAAA AGATTACTTT CATTCTTCTT TCTCCTTTT CTTTATTTCA TTTAATACAA	1800
TTTCTGTGAG TTTTTCACCT GCTTCTGACA AGGTTACATC TTTAATTGGA GCTGCTGCCA	1860
AATTAAAGTG GCCTCCACCG CCTAACTCTT CCATAATCCG TTGTACATTC AGTTTACTAC	1920
GACTTCGAGC TGAGATAGAG ATAAATCCTT GTGTATTCTT CGCAAGAACA AAACGCGTT	1980
CAATACCTGA CATGGCTAAC ATGGCATCTG CTGCCTTACT AATAACAACGT GTATCATAGC	2040
ATTTATGTC CTTAGCCTCT GCTATTAGTA CATCTGAACC TAATTTACGC CCCTGTAAAA	2100
TAAGTTCATT GACCTCACGA TATTCTTCAA AATCTGTCGC AGCGATTTC TGGATAGCAA	2160
TACTATCACT TCCGCGCGTT CTGAGATAGC TAGCAACATC AAATGTCCGA CTAGTTACTC	2220
GCGAGGTGAA ATTTTGTAGTA TCCAACATCA TACCAGCCAT CAAGACACTT GCTTGCATAC	2280
GACTCAAACG ATTTTCTTGA GAATTCTGGA ACTGAATCAA TTCCGTTACC AACTCACTGG	2340
CACTACTTGC ACCACTTTTC ATATAAGTAA TAACCGCATT ATCTGGAAAA TCCTGATCCC	2400
TTCTATGGTG GTCAATAACA ATGGTTTGGG TAAATAAATC ATAAATTTCT TTTGATAATG	2460
TTAAGGCTGT CTTTGAATGG TCTACAAGAA TCAACAAAGA ACGATTGGTC ACCATCCCCA	2520
TTGCATCCTT AACAGACAAC AACTTCGTAA CTCCTTCTTT TTCTATGAAT GAAACAGCTC	2580
GTTCATATC TGGAGACATT TGTTCTTCAT CATAAAGAGC ATAGCTATTT TCAATCACAT	2640
TGCTGGCGAA CAACTGCATA CCTACAGCAG AGCCCAAAGC ATCCATGTCT AAATTTTGT	2700
GACCGACTAC AAAAACCTGA TCTACACTCC GAATCTTATC TGAAATAGCT GTCATCATAG	2760
CGCGCGTACG AGTCCGTGTA CGCTTGATTG AAGCAGCAGA CCCACCACCA AAATAAACTG	2820

235

GATTTTTCGT TTCGTCGTTT TCCTTAACAA CCACCTGGTC GCCACCACGT ACTTCAGCCA	2880
AGTTCAAATT GAGCAAAGCA ACTTTCCCTA TCTCATCATG ATTTCCATCG CCATAAGAAA	2940
ATCCCATACT TAAGGTCAAG GGCAACTGTC TCTGTTTCGA CTCTTCTCTG AAAGCATCAA	3000
TAACAGAAAA TTTATCATTC ATCAAGCCCT CAAGCACCGT GTAGTCAGTA AATAGATAAA	3060
ATCGATCCAT ACTTACCCGA CGAGAAAACA TCATGTGTTT TTCTGAAAAC TCTGATATAA	3120
AATTAGCTAC AAAACTATTG ATTTGACTAA TATCTGACTC AGAAGTTTCA TCCTCCAAAT	3180
CATCATAATT ATCCACAGAG ACAATCCCAA TCACTGGTCT ACTTGTACC AATTCATCTG	3240
TTATGGCTTG TTCCCTGGAT ACATCTACAA AATACAAAAC ACCGGAAGAA GCATCCATAT	3300
GAACAGCATA ACGCTTCTCA CCAAGCTTGG CATAAGTAGA CGGATTTCCCT ACTGAAGCCT	3360
TGATAATCGT TTGAACAGCT TCTAAATCAA AATCACCATC TTCCCTGGTC AAAATCAATT	3420
CAGCATAGGG ATTAAACCACTCAACCTCTC CAGAAGATAA ATTCAATTTC ATAACACCTA	3480
CAGGCATCTG TTCCAATAGA GCTGTCAAAC TTTCTTCCGC TTGGTGGTTT ACATACTGTA	3540
TCTGTTCTAC ATCACTCCTT GTATAATGCA CTCTCAGTTT CTTAAATAAA AAAACATAGC	3600
CTCCTACAAA AAGAAACAAA ATTAAAACCG TCAACAGATT ATTATTAACA AAAATAATGA	3660
AAGTGGATAA GACTCCAAAC GCAATCAATC CTACTAGAAT AGGAAAAATT GGACTTACAT	3720
AAAATTTTTT CATTCAAAC CTCTTGGCAC CCATTATACC ATAATACCCC TCAAAAAGCG	3780
ACTTTTAAAA AGTGTAATCA GTAATTCTAT CAATTATAAG AAAAAGGTAG TTTACAATTC	3840
AGTAAACCTA CCTTTACACA TATTGAAATT AAGATTCTTT AACCTCTAAC AAACCAATTT	3900
CGCCATCCTC ACGACGATAA ATCACATTGG TTGTCTGATC TTCAACATCC ACATAGATAA	3960
AGAAATCATG CCCCAATAAA TCCATTGTGA GAATTGCTTC TTCCAAATCC ATTGGTTTTA	4020
AATCAATTTG TTTTGAACGA ACAACTTTAG ACTGGACAAT ATTTGAATCT TCCACCAAAG	4080
CATCTGTAAT TAATTGACCA GTTGCTACCT TATTTTATT TTTACGCTCG ATTTTGTGTT	4140
TATTTTACG AATCTGACGT TCAATTTTAT CAGTTACAAG GTCAATTGAA CCATACATAT	4200
CTTGAGATAC ATCTTCTGCG CGGAGAGTAA TAGATCCAAG CGGAATCGTT ACTTCCACTT	4260
TAGCCGTTTT TTCAAGATAA ACTTTTAAAGT TAATTCGGGC ATCCAACCTCT TGTCTGGTT	4320
GGAAGTACTT TTCGATCTTT TCGAGTTTAG AACTACATA ATCACGAATT GCTTCTGTTA	4380
CTTCTAGGTT TTCACCACGG ATACTATATT TAATCATATG AGTACCTTCT TTCTAAACAT	4440
TTTTGTTTTT ATGATTTTAT TATAACGCTT TCATTCTATT TTTGCAAATT TTTTCTCAT	4500
CTTACAAGGG AAAATGTTTT TACATCTTA GCACCAGCTT CTCCAACAG TTTCTTAACA	4560

236

CGATTTATAG TTGCTCCTGT AGTATAGATA TCATCTATAA GTAGGATTTT TTTAGGAATA	4620
GTGACTCCAC TTTTAATAAA GAAAGGAAGT TCTGTCCCCA AGCGCTCTGA ACGATTTTAA	4680
GAAGAACTGG CTCTCTCTTC TCTTTTCTCT AATAAATCCA GATACTCAAA GCCTGTGCCC	4740
TCTACCAAGC CCTCAACCTG ATTAAATCCT CTATTAGCAT ATCTATCAGG ACTTAGGGGA	4800
ATTACAACAA ATTGATACTC TTTGTACTTT TTCAACTCCT CACTTAAAAA TGAAGCGAAA	4860
ACTTTTCTTA ACAGGAAGTC TCCATCAAAC TTATACCGAC TGA AAAAATC CTTCATAGCT	4920
TGATTGTAAG TAAAAATCGC TCTATGACTG ACTTCAACTC CCTCTTTACA CCAAAGTTGA	4980
CAATCTTGAC ACTTTGTGTA CAACTCTGTT TTCATACAAT TTGGACAGTT CTCTTCCCCA	5040
ATTCTTTCAA AAGTAGAATC ACAGTCTGAA CAAAGACAAG AGTCATCATT CCTCAGAAGT	5100
AAGAGACTAC TAAAAGTTAA AACAGTCTTC ATAGTCTGCC CACATAACAA GCACTTCATA	5160
GACCAGCCTC CTTATTCATC ATCTGAATTT CCTTAATCGC CTTCTTGATT GAAGCATTTA	5220
ACCCATCATG GAAGAAAAGC AAATCTCCTG TCGGTCTATC CATGCTTCGT CCAACTCGTC	5280
CACCAATCTG AATCAAACCTA GACTTGGTAA ACAAACGATG ATTGGCCTCT ACTACGAAAA	5340
CATCCACACA AGGGAAGGTA ACTCCGCGCT CCAAGATTGT CGTACTGATA AGTATTGTCA	5400
GTTCTCCATC TCGAAAAGCT TGTACTTGCT CTAATCGATC CTCTGTTACA GAAGATACAA	5460
AGCCAATTTT CTCATTGGA AATTGCTCCT GTAAGATTTC TGCTAACTGC TCCCCTTTCT	5520
TAATTTCTGA AGCAAAAATG AGTAACGGAT AAGCTGTCTT TCTCTGCTTC TCAATATAGG	5580
ACTTTAACTT TGGTGACAAA CGATTCTTGT CTAAGTAGCG ATTAAAATCC GATAACCAAA	5640
TTGGTTTTGG AATAATCAAC GGATTTCAT GAAACCGTCT CGGTAAATTC AGTCTTTTAA	5700
GTTCTCCTAA ACGGACCTTT TTATCTAACT CATTGGTCGA AGTCGCTGTT AAAAAGATTC	5760
TCAATCCATT CTCCTTTACA CTATTCTTGA CAGCGTGGTA AAGCATGGGA TTATCAACAT	5820
AAGGAAAAGC ATCTACTTCA TCCACTATCA GCAAATCAAA AGCTTGATAA AACTTCAATA	5880
ACTGATGGGT TGTGCAACA ACTAGTGGTG TTCGAAAATA AGGTTCCGAT TCTCCATGTA	5940
GCAAAGCTAT CCCGCAAGAA AAATCCTGTT GCAGGCGCTT GTACAGCTCC AAACAACAT	6000
CTATGCGAGG ACTAGCCAAA CACACTGCAC CACCCGCATT GATCACTTTA GCCACTACTT	6060
GATAAATCAT TTCTGTCTTT CCAGCTCCTG TTACCGCATG AACTAAGGTT GGCTTTTGCT	6120
TGTCTACTAC TTGAAGCAAT CCCTCTGACA CCTTCTCTTG AAAAGGAGTT AATTGGCCGC	6180
GCCATTTGAG AACATCTTGC TTTGGAAAAT CCTCCTGCGG AAAATAGTAT AAAGTTTGAT	6240
CACTTCTGAC TCGCTTCATC AGCAAGCACT CTCGACAATA GTAAGCACCG ATGGGCAAAT	6300
ACCATTCTTC TAGAATAGTA CTATTACAGC GTTGACAGAA AAGTTTCCCC TTCTCCTTTC	6360

TCATTGCTGG AAGTTTCTCC GCCAACTGAC GTTCTTCTTC TGTTAATTCA TTCTCAGTAA	6420
ATAAACGACC GAGATAATCT AAATTTACTT TCATACTTCT TPATTCGTAA AAAC TAGCAC	6480
TTTAGATGAT TTTT TAGTAC AATTAAATCA TGGAAATTAG GACAATTAAA GAGGACGGTC	6540
AAGTCCAAGA AGAAATCAAA AAATCTCGCT TTATCTGCCA TGCCAAGCGT GTTTATAGCG	6600
AAGAAGAGGC TCGTGACTTC ATTACTGCCA TCAAAAAAGA AACTACAAA GCGACACATA	6660
ACTGCTCTGC CTTCAATTATT GGAGAACGTA GTGAAATTAA ACGTACAAGT GATGATGGTG	6720
AGCCTAGTGG TACTGCTGGT GTTCCCATGC TTGGGGTACT AGAAAATCAC AATCTCACCA	6780
ATGCTCTGTGT GGTCTGTACA CGCTACTTTG GTGGTATTAA ACTAGGCGCT GGAGGACTAA	6840
TTCTGTGCTTA CGCCGGCAGT GTCGCCTTAG CTGTCAAAGA AATTGGTATT ATTGAAATAA	6900
AAGAACAGGC TGGCATTGCT ATTCAAATGT CTTATGCTCA GTACCAAGAG TACAGTAACT	6960
TCCTTAAAGA ACATGGTCTC ATGGAGCTGG ATACAACTT TACAGATCAA GTCGATACGA	7020
TGATTTATGT TGATAAGAA GAAAAGAAA CTATTAAAGC TGCCTTGTG GAGTTTTTTA	7080
ATGGAAAAGT CACTTTAACT GACCAAGGTT TACGAGAGGT TGAAGTTCCT GTAACTTAG	7140
TGTAAACAAT GAATAATACA GCGTTTCGTT GACATTCTCA CACTACTTT AGCGAGCAAA	7200
ATAAAAAGAG GCGTACCAA ATATACTAGA AAATGAAGCA ATTCAAACGA AACCTGATAT	7260
CGTTTTCTCTT CACACCTATT TACTAGAATT AGCTGAACGC AATCACTTGA AAATTAATGA	7320
CTTTGATCTA TGATATATAG AAATGGTATG GATAGCGTTA TACTAAAGAT ATCTTATACA	7380
AAGAGGTATT CATATGTCTA TTTATAACAA CATTACTGAA TTAATCGGTC AAACACCGAT	7440
TGTTAACTT AACAACATCG TGCCAGAAGG TGCTGCAGAC GTCTATATAA AGCTTGAAGC	7500
ATTTAATCCT GGTTCATCTG TAAAAGACCG TATTGCCCTT AGCATGATTG AAAAAGCTGA	7560
ACAAGATGGT ATTCTGAAAC CTGGTTCTAC TATTGTTGAA GCAACAAGTG GAAACACCGG	7620
TATTGGACTT TCATGGGTAG GTGCTGCTAA AGGGTATAAA GTCGTCACTG TTATGCCTGA	7680
AACTATGAGT GTAGAACGAC GTAAAATTAT CCAAGCTTAT GGTGCTGAAC TCGTCCTAAC	7740
TCCTGGTAGC GAGGGAATGA AAGGTGCTAT TGCTAAGGCT CAAGAAATCG CTGCTGAACG	7800
TGATGGTTTC CTTCTCTTC AATTTGACAA TCCAGCTAAT CCAGAAGTAC ACGAAAGAAC	7860
AACAGGAGCT GAGTACTAG CTGCTTTCGG TAAAGATGGA TTAGATGCCT TTGTTGCTGG	7920
AGTAGGTA CT GGTGAACGA TTTCTGCTGT TTCTCATGCA CTCAAATCAG AAAATTCTAA	7980
CATTCAAGTT TTTGCAGTAG AAGCAGATGA ATCTGCTATT CTATCTGGTG AAAACCTGG	8040
TCCTCACAAA ATTCAAGGTA TCTCAGCTGG ATTTATTCCT GATACACTTG ATACTAAAGC	8100

238

CTATGATGGT ATCGTTCGTG TAACATCAGA TGACGCTCTT GCACTCGGAC GTGAAATTGG	8160
TGGAAAAGAA GGCTTCCTTG TAGGGATTTC CTCAGCTGCA GCTATCTACG GAGCCATCGA	8220
GGTTGCCAAA AAATTAGGTA CAGGTAAAAA AGTCCTTGCC CTAGCACCAG ATAACGGTGA	8280
ACGTTATCTC TCTACAGCAC TTTATGAATT GTAACCGTCC AATAACGAAG TCTATTGAAA	8340
AATCTCCAGA CTAGAGAACT CACGGATAGT TCCTAATCTG GAGATTCTT ATTTGCACTT	8400
TTCTTGACA ACTTTAGTCC ATGGTAAATA GGCCTCTAAA ACCTCTTTGT TTACGAGAGT	8460
TTCCACGTTT GGAAGACATT CTAGAAGATA GGATAGATAT TTCTCACTAT TTATAATGGA	8520
TTGAAATAAG ATATGAACAA ATCGATTAGA ACATGATGGT AAAGCGTAAT CCCTTGTTTC	8580
TCAGCTTTCC CAGACAAAAA AGTCCAATAG TAAGTCAGCT GACTATCACT CTCTAGCACC	8640
CTATAAGAAG TTTCATCCGC ATGAAGTAAG GGCTGAGTCA ATAGTCTCTC TCGCAAGAGG	8700
TTATAAAGGG GCTCCAAATA GTATTGACTC GTCTTGATAT GCCAATTAGA GATTTCCTTA	8760
CGTGTGATTG GTAAACCCAT CCTAGCCCAA TCTTCTTCTT GGCGATAATT GGTACCTTC	8820
AGATTAACT TCTGATGGAT GGTGTGAGCG ATAATAGAAG CTGAGCCAAA GTTATGCGCT	8880
AAAGGGGCTT TAGGAATAGG AGCTTTCACA AGCTTATCCA GATGATTATC TTTTACTCGT	8940
TATGGACAAT GCTATATGGC ATAAATCAAG TACCTTAAAG ATTCCGACTA ATATTGGCTT	9000
TGCATTTATT CCTCCATACA CACCAGAGAT GAACCCCAT GAACAAGTGT GGAAAGAGAT	9060
TCGTAAACGT GGATTTAAGA ATAAAGCCTT TCGAACTTG GAAGATGTCA TACAAGGACT	9120
GGAGAAGGAG GTGATAAAGT CCATCGTTAA TCGGAGACGG ACTAGAATGC TTTTGTAAAA	9180
CAGATGAGTA TAAAAAGAAA GTCCTCATTT CAATAGAAAT CACGACTTTC TGATGAATTT	9240
ATAGTAAAT GAAATAAGAA CAGGATAGTC AAATCGATTT CTAACAATGT TTTAGAAGCA	9300
GAGGTGTAAT ATTCTAGTTT AAATCCACTA TATTTGGGGA GTGATAGAAA AGCCCTTCAT	9360
CAGCCAATCT ACTTGTTCAG GTGCGAGAGC TTTGACATCC TTTTCTGTAC TGGACCAAGT	9420
CAGTTTTCG TTCTCAAAGC GTTATATATA TATCCAAAT CCTTGACCAT CCCAGTAAAG	9480
AACTTTAAAG CGGTCTTTAC GTCCACCACA AAAGAGAAAG ACTTGATCGG AGAAAGGATC	9540
CAATTCAAAG TGGGTTTTAA CTACATAGGC TAATGAGTCT ATTCCCTGCC TCATATCTGT	9600
CTTGCCACAA ACAAGGTGAA CTTGACCTAA ATCACTTAGT TGAATTATCA TAGTACAATA	9660
CCTTCCTCC GATAATTATT TTTTATCTGG TATACTGGAA GTTGGGGAAT TAGGATAGAT	9720
ACCTTGTTAT GACGCGCTTA CTATGAATTT GAAGTATAGT CTCCTAAATG CACTTAGCCC	9780
TTATTATAGG GCITTTTGTT TTAATTATTC TAATCGAGTG AGACTGGGGA AAAACAATT	9840
TCAGGAAAAA TCTAAGCCCT ATACAAAAA GGAAGCAATT TGCTTCCTTT CTATTATTAG	9900

239

TTATTCAAGG CTGCTGCCAT TGTAAGCTGCA ACTTCAGCTT CGAAGTCGTT TGCAGCTTTC	9960
TCGATACCTT CACCAACTTC AAAGCGAGCA AACTCAACTA CCGAAGCGTT AACTGATTCA	10020
AGGTATGCCTT CAACTGTCTT GCTGTCATCC ATGATGTAAA CTTGTGCAAG AAGTGTGTAA	10080
GCTTGGTCAA CTTTAGTGTT ATCAAGCATG AAGCGATCCA TTTTACCTGG AATAATTTTG	10140
TCCCAGATTT TTTCTGGTTT GCCTTCTGCA GCCAATTCAG CTTTGATGTC AGCTTCAGCT	10200
TGAGCAATAA CATCATCAGT TAATTGAGCT TTTGATCCAT ACTTCAAGTG TGGAAGAGCT	10260
GGTTTATTAA CCATTGCACG GCTTTCGTTG TCTTGGTCGA TAACGTGATT CAATTGTGCC	10320
AACTCATCTT TAACGAATTG CTCATCCAAT TCTTTGTAAG AAAGAACTGT TGGTTTCATC	10380
GCTGCGATGT GCATTGACAA TTGTTTAGCA AGTGCTTCGT CTCCACCTTC AACAACTGAA	10440
ATAACACCGA TACGTCCACC GTTATGTTGG TATGCTCCAA AGTGTTGTGC GTCTGTTTTT	10500
TCAATCAATG CAAAGCGACG GAATGAGATT TTCTCTCCGA TAGTTGCTGT TGCAGATACG	10560
TATGCAGCTT CAAGAGTTTC ACCTGAAGGC ATTATCAAAG CAAGAGCTTC TTCGTTGTTA	10620
GCAGGTTTTT CTTAGCAAT GACTTTAGCT GTAGTATTTA CCAATTCAAC GAATTGAGCG	10680
TTTTTTTGCAA CGAAGTCAGT TTCAGCGTTT ACTTCAATAA CTGCTGCAAC ATTACCGTTA	10740
ACATAAACAC CAGTCAAACC TTCTGCAGCA ACACGGTCAG CTTTCTTAGC TGCCTTAGCC	10800
ATACCTTTTT CACGAAGCAA TTCAATCGCT TTTTCGATGT CACCGTCTGT TTCTACAAGC	10860
GCTTTTTTAG CGTCCATAAC ACCGGCACCA GATTTTTCAC GCAACTCTTT TACAAGTTTA	10920
GCTGTAATTT CTGCCATTTT AATTCTCCTA TATTTTTTGA AAATAGGAGA GCGCGGCTAA	10980
GCCCCGCCTC CGG	10993

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8411 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CGACGGGGAG GTTTGGCACC TCGATGTCGG CTCGTGCGAT CCTGGGGCTG TAGTCGGTCC	60
CAAGGGTTGG GCTGTTCCGC CATTAAAGCG GCACGCGAGC TGGGTTTACA ACGTCGTGAG	120
ACAGTTCGGT CCCTATCCGT CGCGGGCGTA GGAAATTGA GAGGATCTGC TCCTAGTACG	180
AGAGGACCAG AGTGGACTTA CCGCTGGTGT ACCAGTTGTC TTGCCAAAGG CATCGCTGGG	240

240

TAGCTATGTA	GGGAAGGGAT	AAACGCTGAA	AGCATCTAAG	TGTGAAACCC	ACCTCAAGAT	300
GAGATTTCCC	ATGATTATAT	ATCAGTAAGA	GCCCTGAGAG	ATGATCAGGT	AGATAGGTTA	360
GAAGTGGAAG	TGTGGCGACA	CATGTAGCGG	ACTAATACTA	ATAGCTCGAG	GACTTATCCA	420
AAGTAAGTGA	GAATATGAAA	GCGAACGGTT	TTCTTAAATT	GAATAGATAT	TCAATTTTGA	480
GTAGGTATTA	CTCAGAGTTA	AGTGACGATA	GCCTAGGAGA	TACACCTGTA	CCCATGCCGA	540
ACACAGAAGT	TAAGCCCTAG	AACGCCGGAA	GTAGTTGGGG	GTTGCCCCCT	GTGAGATAGG	600
GAAGTCGCTT	AGCTTTAATC	CGCCATAGCT	CAGTTGGTAG	TAGCGCATGA	CTGTTAATCA	660
TGATGTCGTA	GGTTCGAGTC	CTACTGGCGG	AGTAATCGAT	AAAAGGGGAC	ACAGCTGTGT	720
TCCTCTTTTT	GTATCAATTT	GTATCACCAA	GCATTTTCAT	AAGGAAGTCT	GTTATTTCTT	780
GAGAACTTTC	TTTTTTTCCA	TGTGCAATCC	AAGTTTGGCA	GACACCAAAA	AGTGCATGAG	840
TTAGATAGAT	GCTACTATAT	TCTAATTCAG	TGGTATTTAG	ATTCAGTTGC	ATAAATCGCT	900
TTTGTAAATC	TGTACTAAGC	ATGATATGAA	GTTTATTTTCG	TAAGAAATTT	TGGATTTCTT	960
TAGTCCCAT	TTCAGAAAGA	AGGGCAGCCA	GAAGTGGTTC	TGACTCTAGA	TATTCAAAAA	1020
CTTCTAAAAT	AGCGTCTCTT	TTGTGATGAG	CATGTTTTTG	AAAAATATAT	TCAAATGTAT	1080
GGAATAGCTT	GCTTTGATAG	TGCTCAATCA	TATCATACTT	ATCCTTATAG	TGAGTATAGA	1140
AGCTGGAACG	ACTAATTCGG	GCTTTTTCTA	CTAATTTGAC	AGTAGAAATT	TTATCAAATG	1200
GCTGTTCCAT	CAGTAATTGT	ACCATAGCAT	TTTCAATAGT	TCGCTTTGTT	TTTAAGCGTT	1260
TGTTACTTTC	TTGCATATTT	CCTCCTTGTA	AACAAATTAG	ACTATATGTC	TAAAAATAGA	1320
TTTTTTATCT	TGTAATTTAG	ATTTTTTAAT	GTATAATCTA	TTATATCAAA	ATTTTAGACA	1380
ATATGTTTAA	AAAAGCAGAA	ACTAAGTTTA	AAGAATGGAA	AGCAATTTAA	AAAAAACCAA	1440
CCTTTATTAT	TGTCATGATC	GGGATTTCTC	TTATTCCAGA	TCTGTACAAT	ATCATATTTT	1500
TGTCATCAAT	GTGGGATCCA	TATGGGCAAT	TGCTGACTT	ACCTGTGGCA	GTTGTAAATA	1560
ATGATAAAGA	GGCTTCCTAT	AATGGTAATA	CTATGGCAAT	AGGAAAAGAC	ATGGTGTCCA	1620
ATTTAAAAGA	AAATAAAACC	TTGGATTTTC	ATTTTGTAGA	TGAAGAGGAA	GGAAAGAAGG	1680
GATTGGAAGA	TGGCGATTAC	TATATGGTAG	TGACTTTACC	AAGTGATTTA	TCTGAAAAAA	1740
CAACTACATT	ATCCAATATT	CAATCGACAG	CAGCTTATCA	ATCATTGACA	AGTGAGCAAC	1800
AAACTGAGAT	AAGTGATTCT	GTATCTCAAA	ATTCAACTGA	TAGTATTCAA	TCGGCTCAGT	1860
CAATTGTAGC	TTTAGTACAA	GATTTACAGG	GAAGTTTAGA	AAACTTACAA	AATCAATCTT	1920
CTAATCTTTC	GACTTTAAAA	AATCAATCTA	ATCAAGTATC	ACCTATTACT	TCTACTTCTT	1980
TGATAGGATT	GTCAAGTGA	TTAACAGAGA	TACAAGGAGA	TGTTACTAGC	AAATTAGTTC	2040

241

CTGCCAGTCA	GTCGATTGCA	TCAGGTGTAA	ACGCATATAC	TACAGGTGTT	GATAAAAGTTT	2100
CTCAGGGCGC	AAGTCAACTA	AGTGAAAAAA	ATGCCACCTT	GACAGGTAGT	TTGGATAAAC	2160
TAGTTTCAGG	CTCAAAACACC	TTGACACAAA	AATCTTCTAG	ATTGACAGCA	GGAGTTGGTT	2220
AATTACAATC	AGGATCTGGG	CAATTAGCAG	ACAAATCCAG	TCAGTTACTT	TCAGGTGCTT	2280
CTCCATTAGA	GAATAGAGCT	AATAAATTGG	CAGATGGATC	TGGGAAACTA	GCAGAAGGTG	2340
GAACAAAGTT	AAC TTCTGGA	TTGGAAGATT	TACAGACAGG	ACTTGCTTCT	TTAGGACAAG	2400
GACTAGGTAA	TGCTAGTGAT	CAACTCAAAT	CAGTATCAAC	AGAATCTAAA	AATGCAGAGA	2460
TTTTGTCAAA	TCCACTCAAT	CTTTCAAAAA	CAGACAATGA	TCAAGTTCCT	GTAAATGGAA	2520
TCGCAATAGC	TCCTTATATG	ATATCAGTTG	CTCTTTTTTT	GCAGCAATAT	CAACAAATAT	2580
GATATTTGCG	AAATGSCCTT	CAGGACGTCA	TCCAGAGAGC	CGTTGGGCTT	GGTTGAAATC	2640
TTGAGCTGAA	ATAAATGGTA	TTATAGCTGT	TTTGGCAGGA	ATTTTGGTAT	ATGGAGGAGT	2700
TCAGCTTATT	GGTTTAACTG	CTAATCATGA	GATGAGAATA	TTTATTCTCA	TCATCCTAAC	2760
AAGTTTAGTA	TTCATGTCTA	TGGTGACCAC	TTTAGCAACG	TGGAATAGCC	GTATAGGAGC	2820
TTTTTTCTCA	CTTATTTTGC	TTTACTACA	GTTAGCATCA	AGTGCAGGTA	CTTATCCACT	2880
TGCTTTGACA	AATGATTCTT	TTAGATCTAT	TAATCCCTGG	TTACCAATGA	GCTATTCACT	2940
TTCGGGATTA	CGACAAACAA	TCTCTATCAA	CAAGTCATTT	TCCTAGCTGT	CATACTAGTT	3000
CTATTTACTA	GTTTAGGTAT	GCTAGCCTAT	CAACATAAGA	AAATGGAAGA	AGATTAAAAA	3060
AATCGACCGA	TTAACTGGTC	GATTTTTTAT	CCCTTAGATG	ACTTTCGTCT	GTGATTATAG	3120
ATTCCAAATA	GTAAGAGAGA	AGTAAAGGAA	CAGATTGCTC	CAGTAATAAA	ACCATTGGGA	3180
ATGAAGGAAA	GTGTAATAGT	TCCTTTCCCC	TTGGGAATGT	CAACTTTCAT	AAATCCAGTT	3240
TGAGCTTGTT	TAATTTCTAT	TTTCTTACCA	TCTTGGTAGG	CAGACCAACC	TTTGTCTATA	3300
GGAAATGGTGA	AGAAAATAGA	TGTATCTTGT	TGGACATCAT	ATGTAGCAAA	AACCTTGTTT	3360
TTAGAAGTTG	ATACTGTGAC	AGGTGTCTCT	TTAATTTTTT	GAATTGCCTC	GGTGAAAGTT	3420
TTGGTATCTA	AACGATAGAA	GGTAGGAGAT	TCAAATGATA	CTTGTGAATT	TCCAGGGAAA	3480
CTAACATTGA	TATTGAAAGT	TTTTTTCTCT	TTAGTATATC	CTAGATTAAA	GAAGGAGAAG	3540
ACATTATCAG	TTGTAAAAGT	CTTTTTTTCA	CCATTTACAA	GGATGTCAAC	CTTCTTTTGT	3600
TTATCGTTAG	AAAAGTGAAG	GTTTATGAAA	GAGAGATAAA	CTTGGCTGTT	TTCTGGAAC	3660
TCAATTTGAT	ACTGGATTGC	TGCATCTTCA	TTTGAAGAAC	TTGTGACACT	AATCAAATCA	3720
TTAGTATTTT	CTATTTTTTC	TGTTTTTTCA	TAAGGTATTG	GAGAAAAATA	ATCAAAATTG	3780

242

ACGTTAGCAA GTTGATTTAA AAATGAGGCC TGATTATCCA AGGTATGTTT ATTGAACCTG	3840
ACATCATTTG AAACAGATTG ACTCGCAACT GCAATCGGAA GAGAGTATTG ATTTTCATAT	3900
AGGGTAAGAT TATCTTTTGT ATAGATATCT TTAAAGCCAT ACTTATCAAT AGGACTGTCT	3960
GAGATATTGT ACTGGATACC AAATAAACTA TCAGCCAAAA TACTATTATT TGCATATCGG	4020
AGATTGAGAT TAGTCCCAGA GGATTTAAAA CCAAGTTTAT CTAAAGTAGA GCTTGATGAA	4080
CGATTTTCGAA CAGATGAAAA TTGAGAGATT CCATTGTAGT TGAATTTTCAT ACTGTCATTT	4140
CCTGTCTGAG TTTGTAGTTT TTCAGTACGA GTAAATTGAT TTCCAATATA TGTGAGAAA	4200
GATTCCATAG CTGGGATATC TCGACTATAA GCACTTCGAG AAGCAAATCC CCATTCCTTA	4260
GCAATTCGGT CCATTTGAGA TGAAGCATTT AAATCATTTT CAACCAGTAT AAATAAAGAG	4320
ATTAGAATGG CAAATAGATT CACAGATATA AACTTTTGA TAACTGCAAG GAGTAAAGA	4380
GAATAGACAA CAAAAATTC AAGAGTAAGC AGAATATTCA AATCTGTAA AAAAGAATAA	4440
TGCGATTTTA GATAGATGGT AGCTAAAAAT CCTGCTACTA CAAGAAAAAG CGAAACTAAA	4500
AAATTCCAGA CTTTAAGTTC TTCAGACGC TTTAAGACTT CTGCTGCTGT GTAAATTAAC	4560
AAGGTAGAGA AAATCCAAGC ATAGCGATGT AAAACATGT TTGGAGTATG CATGCCCTGC	4620
CAAAATAAGT CAAGAGCTTC TATGTAAAAG CTTGCAATTA GAAATGCAA GAATATTACA	4680
TATATGAGTT TCACGTGAAA CTAAATAGAT TTCAGCGTAA AAAATAAAT GTCAAAATA	4740
AAGGGAATA GTCCAACAAA AATCATGGG ATGGCCCAT ACTTTGTTGT GTCAAAGGAA	4800
CCAATGAATT GCTTAGCAA GAGATCAAGA TACCAGCTAC TTTCAGTTTG AAATTTGTA	4860
ACTTCAGTCA ATTTTCCCC ATGTGTCTGT AAATCAAATA GAGTGGGAAG AGTCATAATC	4920
AAACTAGCCA TACCAGCTAA AAAGGAGATA ACTATGAAAT CAAGAACAGA TGATTTTCGA	4980
GTCTTAAAGT CCCACGAAAT TTGACAGAGA TACCAGAAAA TAAGAAACAA TACTGTCATA	5040
TATCCAAAAT AATAATTTTG AATAAATAAG ATTGACAGAC TTGTAAAGTA CAATAGGAGT	5100
TTCTTTTCAG TTATCAGTAG ATGTAAACCA GTTATAATTA AAGGAATCAA GATAAAAACA	5160
TCTAGCCAGG TTTTATCTC TAATTGACTG ACAGTGAAAC TCATCAGAGC ATAGGAAGTA	5220
GATAAGGCTA GTTTTAAAA CTGAGGGATA GATTGAAACA ATTTATTCAA ACTAAAAAG	5280
GTTGACAGAC CAATCAATCC AAATTTTAAG AGAGTTGTCA GATAGATAGC ATCTGGCATA	5340
TTCTTTAGAT CAAAAAGTA AACCAGAGGC GCGAGAAAAC TACCCAAGTA ATAAGTAGAT	5400
AGGGCATAGA AGTTTAGCCC TAGACCACTT GTAAAGGTGT AAAACAGATT ACTATTTCCA	5460
TGTAGGATAT TTCGTAAGGC TACATCAAAA ATAACGTATT GATGAAAGCC ATCTCCTAAT	5520
AGAGGAGAGT TGTCGCTATT CCAGTAGATA CTTTGAGATA GATATACTCC AGACATAATC	5580

		242	
ACGTTAGCAA GTTGATTAA AAATGAGGCC	TGATTATCCA AGGTATGTTT ATTGAACTTG	3840	
ACATCATTTGT AACAGATTG ACTCGCAACT	GCAATCGGAA GAGAGTATTG ATTTTCATAT	3900	
AGGGTAAGAT TATCTTTTGT ATAGATATCT	TTAAAGCCAT ACTTATCAAT AGGACTGTCT	3960	
GAGATATTGT ACTGGATACC AAATAAACTA	TCAGCCAAAA TACTATTATT TGCATATCGG	4020	
AGATTGAGAT TAGTCCCAGA GGATTAAAA	CCAAGTTTAT CTAAAGTAGA GCTTGATGAA	4080	
CGATTTCGAA CAGATGAAAA TTGAGAGATT	CCATTGTAGT TGAATTTTCAT ACTGTCATTT	4140	
CCTGTCTGAG TTTGTAGTTT TTCAGTACGA	GTAAATTGAT TTCCAATATA TGTGAGAAA	4200	
GATTCCATAG CTGGGATATC TCGACTATAA	GCACCTCGAG AAGCAAATCC CCATTCCCTTA	4260	
GCAATTCCTG CCATTGAGA TGAAGCATTT	AACTCATTT CAACCAAGTAT AAATAAGAG	4320	
ATTAGAATGG CAAATAGATT CACAGATATA	AACTTTTGA TAACTGCAAG GAGTAAAGA	4380	
GAATAGACAA CCAAAAATTC AAGAGTAAGC	AGAATATTCA AATCTGTAA AAAAGAATAA	4440	
TGCGATTTTA GATAGATGGT AGCTAAAAAT	CCTGCTACTA CAAGAAAAAG CGAAACTAAA	4500	
AAATCCAGA CTTTAAGTTC TTTCAGACGC	TTTAAGACTT CTGCTGCTGT GTAAATTAAC	4560	
AAGGTAGAGA AAATCCAAGC ATAGCGATGT	AAAAACATGT TTGGAGTATG CATGCCTTGC	4620	
CAAAATAAGT CAAGAGCTTC TATGTAAAAG	CTTGCAATTA GAAATGCAAA GAATATTACA	4680	
TATATGAGTT TCACGTGAAA CTTAATAGAT	TTTCAGCGTAA AAAATAAAAT GGTCAAATA	4740	
AAGGAAATA GTCCAACAAA AATCATTGGG	ATGGCCCCAT ACTTTGTGT GTCAAAGGAA	4800	
CCAATGAATT GCTTAGCAAA GAGATCAAGA	TACCAGCTAC TTTCAGTTTG AAATTTTGT	4860	
ACTTCAGTCA ATTTTCCCC ATGTGTCTGT	AAATCAAATA GAGTGGGAAG AGTCATAATC	4920	
AAACTAGCCA TACCAGCTAA AAAGGAGATA	ACTATGAAAT CAAGAACAGA TGATTTTCGA	4980	
GTCTTAAAGT CCCACGAAAT TTGACAGAGA	TACCAGAAAA TAAGAAACAA TACTGTCATA	5040	
TATCCAAAAT AATAATTTTG AATAAATAAG	ATTGACAGAC TTGTAAAGTA CAATAGGAGT	5100	
TTCTTTTCAG TTATCAGTAG ATGTAAACCA	GTTATAATTA AAGGAATCAA GATAAAAACA	5160	
TCTAGCCAGG TTTTATCTC TAATTGACTG	ACAGTGAAAC TCATCAGAGC ATAGGAAGTA	5220	
GATAAGGCTA GTTTTAAAA CTGAGGGATA	GATTGAAACA ATTTATTCAA ACTAAAAAG	5280	
GTTGACAGAC CAATCAATCC AAATTTTAAG	AGAGTTGTCA GATAGATAGC ATCTGGCATA	5340	
TTCTTTAGAT CAAAAAGTA AACCAGAGGC	GCGAGAAAAC TACCCAAGTA ATAACAGAT	5400	
AGGGCATAGA AGTTTAGCCC TAGACCACTT	GTAAAGGTGT AAAACAGATT ACTATTTCCA	5460	
TGTAGGATAT TTCGTAAGGC TACATCAAAA	ATAACGTATT GATGAAAGCC ATCTCCTAAT	5520	
AGAGGAGAGT TGTGCTATT CAGTAGATA	CTTTGAGATA GATATACTCC AGACATAATC	5580	

243

ACTACAGGAA TGATGAAAGA AATAAAATAG GTTCGATATG TTTTAAAAA TGATTTTCATG	5640
TTACCTCGTA GAATGATAGA AAACCTCAGTT GGTAAACCCA ACTGAGTTT GAAGTTTAT	5700
TTAGTCTTTC CAAAGTTCTT TAACTTTTGC TTGTACTTCT GCATTTTCTA GGAATTCATC	5760
GTAGGTTTCA TCGATACGGT CAATGACGCC ATTTTATAGAT AAGACAATGA TATGGTTAGC	5820
CAAAGTTTGA ATAAATTCGT GGTTCATGGCT GGCAAAGATG ATTGATTCTT TAAAGTTT	5880
CAATCCATCA TTCAAGCTTG AGATAGATTC CAAGTCCAAG TGATTGTGTG GATCATCAAG	5940
TACAAGGACA TTTGATTTTA AGAGCATGAG TTTTGAAAGC ATGACACGAA CTTTTCTCC	6000
CCCTGACAAG ACATTTACAG GTTTGTAAAC TTCATCTCCA GAGAAGAGCA TACGGCCGAG	6060
GAAGCCACGT AGGAAAGTAT TGTCATCTTC TTCTTTACTT GCGAATTGAC GCAACCAGTC	6120
AAGAATTGAT TCTCCTCCTG CAAAATCAGC TGAGTTATCT TTTGGTAGGT AAGATTGACT	6180
AGTTGTAAC TCCCACTTGA CAGTTCCTTC ATAGTCAATA TCTCCCATGA TTGCACGAAT	6240
TAATGCAGTC GTTTGAATAT CATTTGTGCC AATAAGTGCT GTCTTATCAT CTGGACGCAA	6300
GATGAAACTA ATATTATCCA AGATAGTTTC ACCATCAATC TTTACAGTTA AATTTTCTAC	6360
TGTCAAGAGA TCATTACCAA TCTCACGTTC CGCTTTAAAG TTGATAAATG GATATTTACG	6420
ACTAGATGGC ACAATCTCTT CTAGCTCAAT CTTATCAAGC ATTCTCTTAC GTGATGTTGC	6480
CTGCCTTGAC TTAGAAGCAT TGGCAGAGAA ACGAGCAACA AATTCTTGCA ATTGTTTAAT	6540
TTTTTCTTCT GCTTTAGCAT TACGGTCTGC TAGCAATTTA GCAGCAAGCT CAGAAGATTC	6600
CTTCCAGAAG TCGTAGTTTC CGACATAGAG TTTGATTTTT CCAAAGTCAA GGTCCGCCAT	6660
GTGAGTACAA ACTTTGTTTA AGAAGTGACG GTCGTGGGAT ACTACGATAA CTGTGTTATC	6720
AAAGTCAATC AAGAAGTCTT CTAACCAAGT AATCGATTGG ATATCCAAAC CGTTAGTAGG	6780
CTCGTCCAAG AGAAGAACAT CTGGTTTACC AAAAAGTGCT TTGGCGAGGA GAACCTTTAC	6840
TTTTTCACCG TTGGCCAATT CGCTCATGTT TTGGTAGTGT AATCTTCTG GAATGTTAG	6900
GTTTGAAGT AGTTGAGAGG CTTCACTCTC TGCTTCCAA CCTCCAAGTT CGGCAAACTC	6960
TCCTTCGAGT TCGGCAGCAC GAACCCCGTC CTCGTCTGAG AAATCTTCCT TCATGTAGAT	7020
AGCATCTTTC TCTTTCATGA TGCTATAAAG TTTTTCATTT CCCATGATAA CGACATCAAT	7080
GGCACGTTCA TCTTCGTAGT CAAAGTGATT TTGACGAAGA ACAGAGAGAC GTTCATCTGG	7140
ACCAAGAGAG ATGTGACCAG TAGTAGGTTT GATATCTCCA GCTAAAATTT TAAAAAGGT	7200
TGATTTTCCG GCACCATTAG CACCGATTAA TCCGTAAGTA TTTCTTCTG TAAATTTGAT	7260
ATTGACATCA TCAAAAAGTT TGCGATCACT AAAACGTAGT GAAACATCAG ATACTGTAAG	7320

244

CAATGTTTTT CTCCTATATG TGTAATATAT TTATTTCTACT AGAAAATACA GAAATATTC	7380
AATTTTTTATT TGTCAATTTT GTGTAAATTA TATTTACAGT ATCCTTTACA CAAATCTGTA	7440
AAAAGCAAGG CTGATTTTATT TTGATAAATT ACGGTTATTT CATTAAAAAA ATGCTATAAT	7500
TGAAAGGACT ATATCGAAGG AGAACAAAAT GACTAAACCC ATTATTTTAA CAGGAGACCG	7560
TCCAACAGGA AAATTGCATA TTGGACATTA TGTGGGAAGT CTCAAAAATC GAGTATTATT	7620
ACAGGAAGAG GATAAGTATG ATATGTTTGT GTTCTTGGCT GACCAACAAG CCTTGACAGA	7680
TCATGCCAAA GATCCTCAAA CCATTGTAGA GTCTATCGGA AATGTGGCTT TGGATTATCT	7740
TGCAGTTGGA TTGGATCCAA ATAAGTCAAC TATTTTTATT CAAAGCCAGA TTCCAGAGTT	7800
GGCTGAGTTG TCTATGTATT ATATGAATCT AGTTTCGTTA GCACGTTTG AGCGAAATCC	7860
AACAGTCAAG ACAGAGATTT CTCAGAAAGG ATTTGGAGAA AGCATTCCGA CAGGATTCTT	7920
GGTCTATCCA ATCGCTCAAG CAGCTGATAT CACAGCTTTC AAGGCTAATT ATGTTCTGT	7980
TGGGACAGAT CAGAAACCAA TGATTGAGCA AACTCGTGAA ATTGTTCTGTT CTTTAAACAA	8040
TGCATATAAC TGTGATGTCT TGGTAGAGCC GGAAGGTATT TATCCAGAAA ATGAGAGAGC	8100
AGGGCGTTTG CCTGGTTTAG ATGGAAATGC TAAAATGTCT AAATCACTAA ATAATGGTAT	8160
TTATTTAGCT GATGATGCGG ATACTTTGCG TAAAAAGTA ATGAGTATGT ATACAGATCC	8220
AGATCATATC CGCGTTGAGG ATCCAGGTAA GATTGAGGGA AATATGGTTT TCCATTATCT	8280
AGATGTTTTT GGTCCGCCAG AAGATGCTCA AGAAATTGCT GATATGAAAG AACGTTATCA	8340
ACGAGGTGGT CTTGGTGATG TGAAGACCAA GCGTTATCTA CTTGAAATAT TAGAACGTGA	8400
ACTGGGTCCG G	8411

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9064 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TGCCGTACTC AAGTACAGCC TGCGCTAAGT TTCCTAGTTT GCTCTTTGAT TTTCATTGAG	60
TATTAGTAAC CAAAATCCGA CCACATAGCC AGCCCCTATG AATATAGCCA TTAAAGCTAG	120
CATGGAATTT AGGAAATTAA AAACCACCGC AGATACAAAG GTTAGCACAA AAACATTAAA	180
AGCAATGGTG TCAGAAGCCA AGACTAGAAT ATAGGGTGTC AACCGATCTA AAGTTTGGGA	240
ATCTAGGAAA AATAAGTGTT TATACATGAT GACCTCCTCT ATGGCTGAAA AGCAAGCCTT	300

245

TTGTTTTTTT ACCCCAAGAC CCTATGTAGA AAAGTGAGCA AAAACGGGAA GGTGCTACA	360
ATATTATTGA TCACATGCAC CGCATAGGAT GGATAAATGC TCTGGGTATA GCGGGTCAAA	420
CCAGCAAAGA TGATTCCAAC TGTGCAAAG ACGAAGATAT CTAACAGACT AGGCAGGCTT	480
GAAAAATGAG GGAGAGCAAA TAAAAAGAA GGAAGAAGCA AATCAAGACC AAATCGCGAA	540
TGCTTAAAGA AAGCATGTTG CAGTAATCCT CTATAAATCA ATTCTTCCAT CAGTGGAACC	600
AGAAAGAACA GGGCTATATA AATACCTAGC TCTGCAAAGT TAGTCCCACT ATAACCAATC	660
AATACAGCCC AACCTTCCGC AGTTGACTGA ACATGTTTAG CTGTCTGAAC GTTAAAGAG	720
ATCTGGAACA CTAGCACTAA TACTGTCAAA ATCGAATACC AAAGCCATTT TTTTCTTGGA	780
ATGCGGAAGA GATAACCATG GCCTGTCTTA ACAAGAACCA CAATCATGAC TCCAATAAAA	840
AGTAAACTCA AGATATTTTG AATCCAGAAT AAATTGCCTA TCTGAGAAGA AAATTGCCAA	900
TAGTTTGGGA CGATAAGCGT CAGCTGAGAA AGACTAAATA CGAAAAATAA GTAAGAGAAG	960
ACTGCACTTA TTTTGAATAG AAGTTGATAC TTTTTCATAG AAATCCTCCC TACTATGACC	1020
TCACCTTGTC AGGCTCTACT GCTGTAAGAT TAAGAAGACA GTTTGTTTTT TTTAAGGCTA	1080
ACCTGACTAC TAGATAATAG ATACATTAAG GCATTAAAGA CAATGAAAAT ATGTCCATAG	1140
AATAAAATCA ACCTCGCATC CAAACCAAGA TAAAGTTTGA TTATCAAAAA GATGAGCAAA	1200
AGAATTGAA ACCATAAGGT TTTTCCAAAA ATAAATTAA AGCGATTTTC AATATCTACT	1260
TCCTTGATTT TTACCGCCAC CCCTTTATTA GCAAGAAGGA AAATCCTGCT TCAACAAAA	1320
CCACTGTAAA GAACAAGCCA CCAATAGAT ACGATAGAGA TTTGTAAAAA TGTCCCTAAA	1380
AGAATATCCA ACACACTACT CAAGAAAATA ACAAAAAATA ATCTGTATTT CATATTAAAT	1440
ACCTCCATTC ATTTATTTCA CTAACAATTT AATAGAGCCT TCTACTCAA TATCCTGTCA	1500
GAAAAGGATA GAAAGCTACT TTTTATAATA CTTCAAGCCC CACATGAGCA GAAGCGTGAT	1560
AAACAAGCAG AGAATACACC TATATAAGCG ATTAGTTGTT GATAGAATTC TGTTCCTGAA	1620
ATACCTCTAT ACAAACAAAT GACAAACATA AATCTGCCA AGCCGATAAA CATAAGTTGA	1680
TTGGTTCTAG GACTAACCAA ATCATCATTT ACTTATATTT AAGAGTATCT CTTTTATTTT	1740
AATGTATGTT AGCACTGAAA AGCAAGACAG GCCAATAATA TTTAAATGA ACAGTAACGG	1800
GGTTAAGTCT CTAAAAAAT TATCTACTGA CACTACAAGA AATACTATAC ATATTATAGT	1860
CGAAACTATC TTTTCTTAT CCATAATTAT TTAATCCTTT CCTAACAAAT CCAGCTTATC	1920
AATCAAGAGC GATTTTAAAC ATAATGTAGC AGCACCCGTT GCAACTTTGA CAAGTTTAGT	1980
ATATCATTGT TTTTAAAAAT TTTTCATCCA AATCTTGAAT TGTCATCGAA ACATCTTGAA	2040

246					
TTGTTAAAAA	ATTTAAAAAG	TAAGCATTAA	AAACATACTT	TCCTCTTTAT	ATTGTATTGA
2100					
TACCAACTTG	TTTGTAGACT	TTTCATCCTG	CTATCACATA	TCATTTTGAC	AGGCGAAACA
2160					
ATATTAAAGA	AACTCCCCTG	TAAATTAAGC	TAGCAAATAC	AGGGGAGAAA	TTTATTTTTT
2220					
AGAGAGTACT	ATCCGTATCC	TTTTTGGAAG	ATTTTGAAAA	TATTTTCTTA	ATTAAGTCAT
2280					
CCATATAAGG	ACCAAATATA	CCAACTACTA	AACCAATAAT	AAAACTTTTA	AAATCCATAA
2340					
TTACCACCAA	CATATTGCTG	CATAGGCTAC	ACCTCCAAGT	ATAGCTCCAC	CTGCAGCACC
2400					
AGTTACACCT	ATTCCTATAG	CAAATGGTCC	CAATAGAAAT	GTCAAACCGT	TGTTGCACAC
2460					
CCATCAATTG	CGCCATATGC	AACCCCTGCT	GCACAACTAA	TTTTTCTTCC	CCAATCAATA
2520					
TCTCCACCTT	CAACGCAAGC	AAGCATTTCA	TTATCCATAA	CTGCAAATTG	TGACATCATT
2580					
TTTGTATCCA	TATAGTGTAT	CACTTTTCAG	TTACGGAACA	AGTTTAATAT	AAAAATTATC
2640					
AAAAAACAT	AGGCAATAAA	GAGAAAAATT	AATTTATCAT	AGATTAGAAA	TAATATGACA
2700					
AAACAATTCA	ATGATGTTAA	TTCAATAGTC	TTTTGTTTTT	TATCGGAGAT	ACTTATGGAT
2760					
AGATAAATAA	GATAGGTTTG	AAAAGCGAAG	AGAATAATAA	AGAATATAGC	CTTCATAAAA
2820					
TTTAGCTTTC	ATTTTTATGA	TGTAGCGGTA	TAGGCTAAAT	ATCCACAAAC	CACTGCTCCT
2880					
CCAATTCCTC	CTATTGCAGC	GCCCCATGGT	CCTAGAAGTC	TCCCATATTT	CACTCCACCC
2940					
GCTGCACAAC	CTAAGCAGC	AACTACAGCT	GCTCCTCCGG	AATTACCTCC	ATAAACCTCA
3000					
CTCAGCATTG	TTTCATTTAT	ATTACAATAA	GTATTCATAC	AAGTCTCCTT	TTATTAATAA
3060					
CCACCCGTTG	CCCCTGTTAC	TCCTGCCCAA	AGATCCACAC	CAAATTTAGC	TCCTATGTAT
3120					
CCACATGCTC	CCATAAATGG	TGCTCCAACA	CCACTCGCAG	CACAAATAGC	TGTCCCTAGC
3180					
CCCCAGCCAC	CAAAAGCAGC	ACCACCACCT	TCTAAGACAT	TAGTTTGCCA	ATTATTCTTG
3240					
CCTCCTTCAA	TACTAGATAA	CATAGTTATA	TCCATTTTAT	GAAATTGTTT	CATAATTTTT
3300					
GTATCCATGA	CAAATACTCT	TTTTTATTTT	TAATTTTTGT	CTTGTGTAA	CTTGGACAAG
3360					
TTTAGTATAT	CATCGTTTTT	TAAAATTTTT	CATCCAGATT	TTGAATAGTC	ATCGAAACGT
3420					
CTTGAATTGC	AAAAATTACA	TTAGACTTCC	TGCAAAACTA	GAATCCTAGT	TCATGATTGA
3480					
TAATACCAGC	ACTCAAATTC	ATTCGTAATC	CGAAGCGTTT	ACGATGACTT	CGATAGGTTG
3540					
TTGAAAACAT	TTTAAACGTT	TTTACTTTGG	CAAAGATGTT	CTCAACCTTG	CTTCTCTCCT
3600					
TAGATAGCGC	ATGGTTACAG	GCTTTATCTT	CAACTGTTAG	CGGTTTGAGT	TTGCTGGATT
3660					
TACGTGAAGT	TTGTGCTTGA	GGATATATCT	TCATGAGCCC	TTGATAACCA	CTGTCAGCCA
3720					
AGATTTTACC	AGCTTGTCGG	ATATTTCTGC	GACTCATTTT	GAACAACTTC	ATATCATGAC
3780					
AATAGTTCAC	AGTGATATCC	AAAGAAACAA	TTCTCCCTTG	ACTTGTGACA	ATCGCTTGAG
3840					

247

TCTTCATAGC GTGAAATTTC TTTTACCAG AATCATTCGC TAATTCCTTTT TTTAGGGCGA	3900
TTGATTTTTTA CTTCCTGTCGC ATCAATCATT ACCGTGTCCT CAGAACTGAG AGGAGTCTT	3960
GAAATCGTAA CACCACCTTG AACAGAGTT ACTTCAACCC ATTGGCTCCG ACGGAGTAAG	4020
TTGCTTTCGT GAACACCAAA ATCAGCCGCA ATTTCTTCAT AAGTGCGGTA TTCTCGCACA	4080
TATTGAAGAG TGGCCATAAG AAGGTCTTCT AGGCTTAATT TAGGTTTTCG TCCACCTTTT	4140
GCGTGTTTAA GTTGATAAGC TGTTTTAAAT ACAGCTAGCA TCTCTTCAAA AGTCGTGCGC	4200
TGAACACCAA CAAGACGCTT AAATCGTGCA TCAGTTAGTT GTTTACTTGC TTCATAATTC	4260
ATAGAACTAT AGTAAATGA AATAAGAACA GGATAAATCG ATCAGGACAG TCAAATCGAT	4320
TTCTAACAAT GTTTTAGAAG TAGAGGCGTA CTATTCTAGT TTCAATCTAC TATACTATAC	4380
CATATTTTGT TTCGCAGGGA ATCTATTATA AAAGGGTAAAG TATTGCAAAA ACACTTACCC	4440
TTTTCTTTTA TACTTCATTA AGCTCTACTT TTTATAATAC TTCAAGCCCC ACATGAGCAG	4500
AAGCATGATG ATTAAGCAGA GAACAGCGCC AATATAAGCG ATTATTTGTT GGTAGGATTC	4560
TCCTGCTGTG ATACCTCTAT ACAAACAAAT AATAGACATA AAACCTGTCA AGCCGATGAA	4620
CATAAGTTGA TTGGTTCTAG GACTAACCAA ATCATCATCT TCAAACCTCTC TTATCCTCAT	4680
TTCCCTAGTG AGATAAACAG TAACCAAAAT AGAAGCCAAG TTAATAACTA CTAAAAGAAA	4740
TTGGAAAAC ACGGAAAAAT TTA AAAACTG ACGAGATAGA AATAGATAAG TAGAAACAAG	4800
CAAGGGCAAC TGACCTAAGA ACAATCTCGC AAGGAAGATG TTCCGTTTTT TAGCAAGAAA	4860
AGTTTTTCAT TCTTTTCTCC TTTCTTTTTA TTGATAGCAA AATAGATCAT AACTGCAATC	4920
ACATAGGCTA TGGTATAAAA TAGCTGATAC CAAGCACTCT CCCTAAGCGG ATATAGAAAAG	4980
ATGGACATGA TTAGATACAG AACGAAAATA ATCAGTATTT TTTTCTTCAT AAGATTTCTT	5040
CCTAAATGTG CGATTTATCT TAGTTGAGCA AGAACATTTA CACTGCTAGT ATAGCACTTA	5100
TTTTGACCTT GGATCACTCA AATCATAAAT GGTCATCAA ACCTCTTGAA TTGTAAAAAT	5160
TAAAAAAGCA AGCATGAAAA ACATACTTTC CTCTTTATAT TGTATTGATA CCAACTTGTT	5220
TGTAGACTTT TCATCCTGCT ATCACATATC ATTTTGACAG GCGAAACAAT ATTAAAGAAA	5280
CTCCCCTGTA AATTAAGCTA GCAAATACAG GGGAGAAATT TATTTTITAG AGAGTACTAT	5340
CCGTATCCTT TTTGGAAGAT TTTGAAAATA TTTTCTAAT TAAGTCATCC ATATAAGGAC	5400
CAAATATACC AACTACTAAA CCAATAATAA AACTTTTAAA ATCCATAATT ACCACCAACA	5460
TGTTGCTGCA TAGGCTACAC CTCCAAGTAT AGCTCCACCC GCAGCACCAG TTGCTGCACC	5520
TTGCCATGTT CCTGTTTAA TGCCTAGTTG AAGACCTCTT GCTGCTCCTC CTCCAACACC	5580

248

TGCTTTGGCA AAATCTCCCC AATTGCATCC GCCACCTTCA ACGCAAGCAA GCATTTTCAGT	5640
ATCCATAACA GAAAATTGTG ACATCATTTT TGTATCCATG ACAAATACTC CTTTTTTAAA	5700
AAACTAAAAT AAATCAGAAT AGAATCCTCA TAATTTTACT ATAAGTCTTA CCAACTTAGT	5760
CCCAATTTAT CACCAACCAT ACCTCCTAAG CATGTTAATC CACCCCCAAT TGCACCAATG	5820
TGTGCTCCAA CAAATGCACC AGCAAGTCCA GCTACTCCTA AAGTGGCCAA ACCTGCTCCA	5880
GTTCCACCAG TTATAATTCC CGTAGTGACT CCTGTAATCA GTGCATTTTG ACAATCAGTG	5940
GAGCTATACC CCCCTTCAAC TTTCGCAAGC ATTTTCAGTAT CCATAACCTC TAACTGTGAC	6000
AACATTTTGT TATTCATGAT GAATACCTCC TTTTATTTT CAATTTGTGA CCAAAGTCTT	6060
AAATPCAATA AACAAATAGA TTTTTTATAG TATCTTTTGT ATTTTCTTAA AAAAGTATAT	6120
ACGTCTACTA TCTTCTTAAA GGTAGCAGTA CCTATTTTTT AGTCTAAGAT TTCAATAATC	6180
TTGAGTATCT AAAATATCTT AATTTCTGTA TTCTCCTTGC AATAAAAAGT TTTACTATAC	6240
TATTTATTAA CTTGCAGAAA GCAAAAAATA TTAGTAAATA ATAGTTTATA GTTAAGTTTT	6300
TTATTCCTAC CAATCCATCA ACTAAGTAAA GCATCAACGA TTACATAAAC GATTGATAAT	6360
ATAATTAAAA TTTTGCTAAC TATCTTATTC TCATCATTTCT TAGATAACTT TGATATTTTG	6420
TAAATAAGTA AATAAGACAG TAAATTAATA GCGATAATAA TACTATATTT AAGAATCATA	6480
ATCTTACAAA GAGGACATAA TTCCTGAACC TACACAAAATA AGTGTGCTG CTCCCCCAGT	6540
TATCGGACCA GTCGCAGCAG CTAATAGTAC TGCTCCAATA CAACCACCGA TGCAGATCC	6600
TAAATTGCCT CTTCTCCAC TAACTATTTT GAGTTCTTCA TTATCCATAA CAGAAAATTG	6660
TTCCATCATT TTGTATTCA TGACAAATAC TCCTTTTTTC TTTTTTTATT TTTGTCTTGT	6720
TGTAACCTTG ATAAGTTTAG TATATCATCG TTTTTTAAAA TTTTTCATCC AGATCTTGAA	6780
TTGTATCGA AACGTCTTGA ATTAGCTTTT TTATTTCAAG CCACCTCTAA ATGTTTAAAA	6840
AAAATAATTT CTAATCACTT TTTTACCATT CAGGAAGTTT TAATGACTAT TCAAGATTTC	6900
ATAAAATATG AACTTAGTTT TATGACATAA TAGACCTATC CACTATATGA AAGGAATTGC	6960
CAATGACTTC TTATAAACGT ACATTTGTTC CTCAAATAGA TGCGAGAGAC TGTGGTCTCG	7020
CTGCCTTAGC CTCGATTGCT AAATCTATG GTTCAGATTT TTCTCTAGCT CACTTGAGAG	7080
AACTTGCAA GACCAATAAA GAAGGGACGA CTGCTCTTGG CATGTGAAAA GCCGCTGATG	7140
AAATGGGCTT TGAAACAAGA CCTGTTCAAG CAGATAAAAC GCTCTTTGAC ATGAGTGATG	7200
TCCCCTATCC ATTTATCGTT CACGTTAACA AAGAAGGAAA ACTCCAACAT TACTATGTTG	7260
TCTATCAAAC AAAGAAAGAC TATCTGATTA TTGGTGATCC TGACCCCTCT GTAAAAATCA	7320
CTAAATGTC AAAAGAACGC TTTTCTATG AATGGACTGG AGTAGCTATT TTTCTAGCTA	7380

249

CCAAACCCAG CTATCAACCC CATAAAGATA AAAAGAATGG TCTACTAAGC AAGCTTCCTT 7440
CCTCTGATTT TCAAACAAAA ATCTCTCATT GCTTACATTG TTCTCTCAAG CTTATTGGTC 7500
ACTATTATCA ATATAGGTGG TTCTTACTAT CTCCAAGGAA TCTTGATGA ATACATTCCA 7560
AATCAGATGA AATCAACTTT AGGAATCATC TCAGTTGGTC TGGTTATCAC CTATATCCTC 7620
CAACAAGTCA TGAGCTTCTC CAGAGATTAT CTCCTAACCG TTCTGAGTCA GAGATTAAGT 7680
ATTGATGTGA TTTTATCCTA TATTCGCCAT ATTTTGAAC TTCCCATGTC TTTCTTTGCG 7740
ACACGTCGTA CAGGAGAAAT CATTTCACGA TTCACAGATG CTAACCTCTAT TATAGATGCC 7800
TTGGCTTCTA CCATTCTTTC TCTTTTCTG GATGTTTCTA TTCTGATTCT TGTAGGAGGC 7860
GTCTTACTGG CACAAAACCC TAATCTCTTC CTTCTTCTC TTATTTCCAT TCCTATATAC 7920
ATGTTTCATCA TCTTTTCTTT TATGAAACCT TTCGAAAAA TGAACCATGA TGTATGCAA 7980
AGTAATTCTA TGGTTAGCTC TGCCATTATC GAAGATATCA ACGGGATTGA AACTATAAAG 8040
TCGCTCACGA GTGAAGAAAA TCGCTATCAA AATATAGACA GCGAATTTGT AGATTATTTG 8100
GAAAAATCCT TTAAGCTCAG TAAATATTCT ATTTTACAAA CGAGTTTAAA GCAGGGAACA 8160
AAATTAGTTC TGAATATCCT TATCCTATGG TTGGCGCTC AATTAGTCAT GTCAAGTAAA 8220
ATTTCTATCG GTCAGCTGAT TACCTTTAAC ACACTTTTTT CTTACTTTAC AACTCCTATG 8280
GAAAATATTA TCAACCTCCA AACCAAATC CAATCTGCGA AGGTCGCTAA TAACCGTTG 8340
AACGAAGTCT ATCTAGTCGA ATCTGAATTT CAAGTTCAAG AAAACCCGTG TCATTACAT 8400
TTTTTGATGG GCGATATTGA ATTTGATGAC CTTTCTTATA AGTATGGTTT TGGATGAGAT 8460
ACCTTAACAG ATATTAATCT CACGATTAAA CAAGGAGATA AGGTTAGCCT AGTTGGAGTT 8520
AGTGGTTCTG GTAAAACAAC TTTAGCCAAA ATGATTGTCA ATTTCTTTGA ACCCTACAAA 8580
GGGCATATTT CCATCAATCA TCAGGATATT AAAACATTG ATAAAAAAGT CTTGCGCCGT 8640
CATATTAAIT ACCTACCCCA ACAAGCCTAT ATCTTTAATG GCTCTATTTT GGAAAACCTA 8700
ACCTTGGGCG GTAATCATAT GATTAGTCAA GAAGATATTC TAAAAGCTTG TGAAGTAGCT 8760
GAAATCCGTC AAGACATTGA AAGAATGCCT ATGGGCTATC AAACCTAGCT CTCTGATGGA 8820
GCTGGTCTAT CAGGAGGACA GAAGCAACGA ATCGCTCTCG CTCGTGCTCT TTAACTAAA 8880
TCTCCTGTTT TAATACTAGA TGAAGCTACT AGCGGTCTTG ATGTCTTGAC TGAGAAAAAG 8940
GTTATAGATA ATCTTATGTC TCTAACTGAT AAAACCATTC TCTTTGTAGC CCATCGTCTC 9000
AGTATAGCCG AACGAACCAA CCGTGTCATT GTTCTTGACC AGGGGAAAAAT CATTGAAGTT 9060
GGTA 9064

(2) INFORMATION FOR SEQ ID NO: 18: 250

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7780 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
CTCCATTTTT TTGATTTTCAT AAATAAACAA CCTCTCTGTT AATTTTGTAT AATTATAACG      60
ATATCCAAAGT TACTTGTCAG GTGTTTTTTA AATTTTATAT TCAAAAATAT TTTTCGTTTC      120
AAAAAAAGGA GCCATCAGTT GATTTCAAGC TCCCTTTTAT ACAGAATTAA ACTATTTTAT      180
AGTTTCGACAA TCTTACCTGT TTCAAAGTAG ACAACCCATT CACAGATATT TTTAGCATAG      240
TCACCGATAC GCTCCAAGTA GGAATAACT TGGAAATAAT CACGACCCGT AACAAATGGCT      300
TCTGGATTTT TCTTAATCTC TTCAGTCGCA AGGTCACGGA TAGTTTCAAA ATAGTGGTTA      360
ATTTGCTCAT CCATGGAGGC CACCCGGTAT GCGTCGTCAA CAGAACCATT AAGATAAAGA      420
TCAAGTGCTG CTTCCACAAC GCTTTTAACT TCACGTCCCA TTTTTTAAAT TTCTTCCTCT      480
ACAGCTGGAA TCGCTCTTC CCCCTTCATA CGGATGGTTG CCTGGGCAAT GGCTACAGCG      540
TGATCCCCCA TACGCTCCAC ATCTGATACA GCCTTAAGGA CAGTCAAGAC TGTACGCAAA      600
TCTTGAGAGA CTGGTTGTTG GAGTGCATC ATTTCAAATG ATTTCTTTTC CAGTTTCACT      660
TCGTATTCAT TTAATTCTGC ATCATCTTCG ATGACCTCTT TTGCCAGGTC ACGGTCATGC      720
GTGACAAAAG CACGTACCGT ACGATTGATT TGTGAGAGCA CTTCTTGTC CATAGCGTAG      780
AACTGGTTAT GTAATTCTCT TAAATCTTCT TCAAATTGAG ATCGTAACAT CTTTCATCTC      840
CTTATCCAAA TTTTCCTGTA ATATAGTCTT CCGTTTCCTT GTGTTGGGGA TCAAGGAACA      900
TCTGCTTGGT ATCATTAAT TCAATCAAAT CTCCATCTAG GAAAAATCCT GTCTTATCAG      960
AGATACGTGA AGCTTGCTGC ATGGAACGGG TTACCAGAAG CATGGTGTAC TTGTCTTTTA      1020
GACCATACAA GGTTCCTCA ATTTTACCAG CTGAAATCGG ATCCAAAGCC GAAGTTGGCT      1080
CATCCAAGAG GATGATTTTA GGAAGTAGTG CCAAGACACG GGCCACGCAG ACACGCTGCT      1140
GTTGACCACC TGACAATCCA ATAGCTGAAT CATATAGACG ATCCTTGACC TCATCCCAGA      1200
TAGAGGCACC TTGCAAGGCT TTTTCTACGG CTTTATCCAG AACCTGCTTA TCCTTAATTC      1260
CATTGATACG AAGCCCGTAG ACAACATCTT CATAGATAGT CATAGGGAAA GGATTAGGTT      1320
GTTGGAAAAC CATTCGATT TCCTTACGTA ATTCAACCGT ATCTGTACGC GGAAGTGTAGA      1380
TGTGTGACAC ATTGTACACC ACGGATCCAG TTGTGGTCAC CTCTGGATTG AGATCTCCCA      1440
```

251

TGCGGTTGAG AGACTTGAGG AGGGTTGACT TCCCTGATCC AGATCGACCA ATCAAGGCTG	1500
TAATTTCTTT AGGTTGGAAA GATAGGGAAA CACTATTCAA AGCCTTCTTT TTATTATAAT	1560
AAACGGACAG GTCTGATACC TGTAATAATCG CATCTGTCAT ACGGTTTCCT TTCTAACCAA	1620
AGTGACCAGA TACATAGTCA TTGGTGGACT GTAGCTTGGC ATTTTGGAAA ATAGTTGCAG	1680
TCTTGTGATA CTCAATCAAA TCACCCAAGT AAAAGAAGCC TGTATAGTCA CTTGCACGAG	1740
CAGCCTGCTG CATATTATGC GTTACAATGA TGATGGTAAA GTTTTCTTTG AGCTCAAACA	1800
TGGTCTCTTC TAGTTGCATG GTCGCAATCG GATCCAAGGC TGAGGCTGGC TCATCCATTA	1860
AGAGGATATC TGGCTTAACA GAGATGGCAC GAGCGATACA GAGACGTTGT TGCTGACCAC	1920
CTGATAAGGT CAAGGCTGAC TTGTGGAGAT CGTCTTTAAC CTGATCCCAG AGGGCAGCCT	1980
GACGAAGGGA GGTTCCTACG ATTCATCTA GGACTTGCTT ATCCTTAACT CCAGCACGTT	2040
CATGCGCAAA GGTAATATTA CGGTAAATTG ACTTAGCAAA TGGATTGGGA CGTTGAAAAA	2100
CCATTCCAAT GTGTTTACGC ATTCATAAA CGTTGATTTC TGGACGGTTG ACATCAATTC	2160
CACGATAGAG AATCTGCCCA GTTACTTTAG CAATATCAAT AGTATCATTC ATGCGATTGA	2220
GACTGCGTAA GTAGGTAGAT TTCCCCGATC CCGACGGGCC AATCAAAGCT GTAATTTTAT	2280
TTCTTTCAAA TTGCATATCA ATCCCCTTAA TGGATTCATT TTTACCATAG TAAACATGGA	2340
CATCCTTAGT AGAAAGGGCT ACTTTTCTT CAGGAAAGGT AAGGATATGC TTCTCATCCC	2400
AGTTATATGT TGACATGGCT TCTCCTTTAG GCAGCGGTTA ATTTCTTGTG TAGATAGCTT	2460
CCGAAGTTAC GAGCTCCAAA GTTAAAAATC AGGATAAAGA TCAGGAGCAC AGCGGCAGAA	2520
CCTGCTGATA CAATGGTTCC ATCTGGAATA GTGCCTTCAC TATTGACTTT CCAGATATGG	2580
ACAGCCAAGG TTTCTGCTTG ACGGAAGATA GAGATGGGGC TAGTCACACT GAGGATATTC	2640
CAGTTAGACC AGTCAAGAGC TGGCGCCGAT TGCCCTGCTG TATAGATCAG AGCTGCAGCT	2700
TCGCCAAAGA TACGACCAGA TGCCAAGACG ACACCCGTTA CAATACCTGG AAGCGCTTCC	2760
GGAATAACAA CATGAACCAC TGTCTCCAG CGAGAAATCC CAAGAGCCAG ACCAGCCTCA	2820
CGTTGGGTAT GGTGAACGTG TTTCAAAC TACTCTACAT TACGCGTCAT CTGAGGCAAG	2880
TTAAAGACTG TCAAGGCCAA GGCACCTGAA ATGATTGAAA ATCCATACTC AAAGTGGACT	2940
ACAAAGATCA AGTAACAAA GAGACCCACC ACCACTGATG GTAAAGAGGA CAAAATTTCA	3000
ATACAAGTCC GCACAAAGTT GGTAAACAGGA CCTTTTTTAG CATATTCAGC CAAGTAAATC	3060
CCAGCTCCCA TAGAAAGAGG TACAGAAATA ATCAAGGTAA TGACCAATAG GAAAAAGGAA	3120
TTGTAAAGCT GAATGCCAAT CCCACCACCT GCTTGAAAAG CAGAAGACCT TCCAGTCAAG	3180

252

AAAGACCAAG AGATATGGGG CAAGCCCCGA ACCAAGATAT AGAGAATCAA GGAAGCCAAG	3240
ATTGTCACAA TGATGCTAGC AATCGTATAG AGGACAGCTG TTGCAAGTTT ATCTAATTTC	3300
TTAGCGCGCA TAATTTTCTT TCTCTCTTC TTTCGTAATC AATTTAATCA CACTGTTAAA	3360
AACTAAGCTC ATCAAGAGCA GTACCAAGGC CAGTGACCAG AGAACATTAT TATTTACAGT	3420
TCCCATGACA GTGTTCCCAA TTCCCATAGT TAATATAGAA GTTAAAGTTG CAGCTGGTGT	3480
GGTCAAGGAA GTTGGGATAA CAGCTGAGTT TCCGACAACC ATCTGGATAG CTAGAGCCTC	3540
ACCAAAGGCA CGCGCCATCC CAAAGACCAC TGCAGTGAAA ATACCAGAAC GGGCCGCTT	3600
CAAGATCACA CGCCAGATAG TCTGCCAGCG AGTGGCTCCC ATAGCGAAAC TGGCTTCACG	3660
ATAATAACGA GGAACCGCAC GCAAGCTATC CGTTGTCATA AAGGTTACGG TCGGCAAAAT	3720
CATGACAAAG AGGACGGAAA TCCCTGACAA AATCCCAAAA CCAGTCCCAC CAAAGACACT	3780
GCGAACAAAG GGAACGACGA CTGCAAGCC AATAAATCCG TACACTACTG AAGGAATCCC	3840
AACCAGGAGT TCAATAGCTG GTTGCAAAAT CTTCGCCCTT TTTGGTGATA CTTCGGTCAT	3900
AAAACTGCT GCACCAATAG CAAAGGGTGT TGGGATAAGG GCTGAGAGAA TGGTAACGAT	3960
AAAGGAACCC AAAATCATAG GAAGGGCACC AAATTCTTTA CTAGAAGGAT TCCAAGTTCC	4020
TCCCAAAAGA AAGTCAAAGA TATTACACCC ATTGACAAAAG AAGGTCGACA AGCCTTTTGT	4080
CGCTACGAAA ACCAAAATCA TGGCCACAAG GATGACTATC AAAGAAAGAC AGGCAAAGGT	4140
CAAACCTTTT CCTAATTCTT CCAGACGAGA ATTCTTTGAT GGAAGCAACA TTTTCTTAGC	4200
TAATCTTCTT TGATTCATTA TTGTCTCCCT TCCAACACTG TCACAGTTCC GGCAGCATCT	4260
TTTTCAACCT TCATTTCTT AATCGGAATA TACTTCAATC CTTTGACAAT CCCTTCTTGG	4320
GTCTCATCCG AGAGAACAAA ATTGAGAAAT TCTGCAGCCA ACTCATTTGG CTGCCCAAT	4380
GTATACATAT GCTCATAAGA CCACAAGGGC CAATTATTGC TACTTATATT TTCTGGACTT	4440
AAGTCATAGC CATTCAACTT CATGCTTTTG ACCGAATCAT CTATATAGGT AAGAGATAAA	4500
TAAGAGATAG CTCCTGGACT TTTTGATACG ATTGATTTTA CCGCTCCATT TGAATCCTGC	4560
TCCTGACTTT GCATGGCAGA CTGACCTTCC ATAATGACAG TATCAAAGGT AGCACGAGAG	4620
CCAGAGCCGG CTGCCCCGATT GATAACAGAG ATGGGTAAGT CCTTACCACC AACCTCTTTC	4680
CAATTGGTTA CCTCACCTAT GAAGATTTGA CGAAGTTGCT CTGTCGTTAG GTTATCAACA	4740
TCAACCTCCT TATTGACAAT CAGAGCCAAG CCAGCTACCG CGACCTTGTG GTCAACAAGA	4800
GCAGAAGCAT CAATTCCGTC TTTTTCCTCA GCAAATACAT CTGAGTTTCC TATATCAACT	4860
GCCCCAGACT GAACCTGGGA CAAGCCTGTA CCAGAACCCT CCCCTTGGAC ATTGACCGTT	4920
TTTCCAACAT GGATCGTGCC AAATTCATCT GCCGCTACTT CAACCAAGGG TTGCAAGGCA	4980

253

GTTGAGCCAA CAGCCGTTAT GGATTCTCCA CGATCAATCC AGCTAGCACA GCCTACTAAA	5040
CAAGCCGTCA GCCAAAAAGC GATAAGAGAC AGAGCAAGCT TTTTCTTTT TTTCAGTGT	5100
TTTCTCCTCG AAAATAATTA TGAATACTGT GAATTTTTTA AGTAGTCTT TATGAGTTGA	5160
CGCATGAATT CTACCAAAT TTCTGCGCAA TTGATTATTT ATATAATATA GGCTATATTA	5220
CTCTTTCCTA ACCTCCTTTT TTCATATGTG GATAAAATCT CTGTCTATC CCTTCCCCCA	5280
TTGTCACCCA TTATAGTCAT TTCGTGTCTC TTTTCCCCCT TTTAATGCA AGGGAAATTA	5340
CTCTCCTTAG ATGATAATCC AAAAGCTAGA AAGGTATCTC AAACCTCTCT ACTCTCCCAG	5400
ACTAGTTTAC AACTAAAAG AAAAGATTCT ATTTTATGAG AAATCTAGTT TACAAGCGGT	5460
AAGAACGCTA ATAATAAAC TTCTGTACT CTTTGAAT CTCTCAAAC CAGTGTTTTG	5520
AGCTATCTAT GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAG TAAACTACA	5580
TGTAATGGCA ATCAAGATAT CAAGAATCAT CCTACTAAA AAATCCATAC TTCACTATA	5640
ACATAGAATA AGATATTTGA CTAGCATTTT CATTTGAATC TGAGGCCTT TGGAAAATAA	5700
TTTTTCAAAA CATTTCCAGT AACCTTTGCA AAGCCCAAGC CATTCCTTT AACCAAACT	5760
TGGTACCAAC CATTTGGCAG ACTTCTGCC AGCTGAACGG TTTCTCCAGC CGCATACTTG	5820
ACAAACGCTT CTTGGCCAAT TTCAACCGAC TGTCGACCT GACTCGGTT CAAGGCTAAA	5880
CCAAGAGCGA AACTGGGCTC AAAGCGTTT TTCTTAAAAG TACCCAGATG CAGTCCATTG	5940
CGAGCAATCT TGAGCTTCCA TAAATCTGGC AAAAGTTCTG GCAAGAGATA AAGCTGGTCT	6000
CCAAAAATCT GCAAGATACC CGGTAGATTG ACCTTCAAAT GTTTTGGGC AAATTCCTGC	6060
CACAAGGCAA CTTGTTCACG GCTGAGGTTA CTCTTACTTG CCTTAAATTT AGGAGCTGGA	6120
TTGTTACCCT TAACTGTAG ATGGGCAACA AACTGACCCT CTCCTTAAA CTGATGAGGA	6180
TACATCCGAG CCGTTTCTGG CAGGTCAATA CCAGCTACCA TTCCATTGAT ATGCTCTACT	6240
GGCAACAAGT CAAAATCATA CTCITCCAGC AACCAATTGA CAATCTCTTC GTTTTCTCG	6300
GGTGCCAGG TACAGGTCGA ATAAACCAGA TGACCACCTT CAGCTAACAT GGTCACTGCA	6360
TCCTCCAGAA TTTCTCTTG CAAGCTAGCA CATTGACTCG GATAATCTAA GTCCTAATAG	6420
TCCATAGCAT CAGGTGCTT ACGAAACATT CCTTACCAG AGCAAGGGGC ATCAAGAAGC	6480
ATTAAGTCAA AATAGCCTTT AAAGACCTTG ACCAAGCGGT CGGCAGATTC ATTGGTCACC	6540
ACGACATTTG TCGCTCCAAA ACGCTCCATG TTTTCAACCA AAATCTTAGC CCGTTTGCTT	6600
GAAATTTTCT TGGAAACAAG TAGCCCTCC CTGCTAGAT AGGCTGCCAG TTGAGTTGAT	6660
TTGCCCCCG GTGCAGCAGC CAAGTCCAAG ACCTTCATAC CAGGACTGGG TTGGGCTACT	6720

254

TGAGCCACCA TTTGAGCAGC AGGTTCTTGC GAATAAACTA AACCTGTAGC ATGCTCAGGC	6780
GATTTCCCTG AAACCTTCCC ATAGTGGCCC CAAGGGGTTT GAGTAATGGC ATCAGAAAAG	6840
GAAAGTGTCT CTTCTTTTAA GGGATTGACC CGAAAGGCCG AAACCGCTTC CTCCTCAAAA	6900
GAGGCAAGAA AATCTCTTGC CTCATCTCCT AGTATCTCTT TATATTTTTC AACAAATCCT	6960
TCTGGAAATT GCATTTAAGT TCTTTTCCTT TCGTAAATAT AGGACTGAAT TTCCTCCTGC	7020
ATCTCAAGAG GCACCATCAT GACCGGCTGT CTGGTTTGAA AATCAGGAGC TTCACCAAAA	7080
AGGGTCACAA CCCGATAGCC CAGACTTTC CCTAAAAATAC TAGCTGCGGC ATAATCCCAT	7140
GGTTGCAGAT AAGTGAGATA GGTCAACAAA CGCCCTGACA AAATCTTGGC AAAACTAATG	7200
GCCGCACTTC CATAGACACG AACACCAAGA ACCGCTCGGC TCAAATCAGC CAGCCCCCAT	7260
TCATTGGTTT CCAGCATACC ACTATTCCTT GCAATGAGAA AATCTCCAAG TGGTTTAGTT	7320
TTAAAAGGAG CTAGGGACCT ATCATTTAGA CAAACTGGAA ATTCCCCACC ACCGTGGTAA	7380
CAATCCCCTT TGACCACATC ATAAATCAGA CCAAAGTGC CCTGACCATT TTCAAAATAA	7440
GCCATCATAA CAGCAAAATC TTCCTGCTGG GCTACAAAAT TATTGGTACC ATCAATGGGA	7500
TCAATGACCC AAACCTTGCC CTCTTGAACC GAGGCTCGCA GACAACCTTC TTCAGCACAA	7560
ATCTTATCCT CAGGATAACG GGACAAAATC TCACCAACCA AGAGTTCCTG AACTTCTTTG	7620
TCCAGTCTGG TCACCAAATC TGTGGGAGAG GACTTGGTTT CAACACGCAA GTCTTCCTGC	7680
ATATGGTCAA GAATGTACTG ACCTGCTTTC TTAACAAGCT CTTTAGCAAA TTCAAATTTA	7740
CTTTCCAAGA GAAATCTTTC CTTCCTTTT TTCTTTGGG	7780

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4820 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GTAATGATAT AGGAACACCA GGTGACCTGA TGGGACGTCG TAAGCCTATG AACTACTAGC	60
TGCTAAAGGC TTAAAGATG GTATGGTACC ATATATCTCA AACCAATACG AAGAAGAAGC	120
CAAACAAAAG GGCAAGACAA TCAATCTCTA CGGTAAAACA AGAGGTTTGG TTACAGATGA	180
CTTGGTTTTG GAAAAGGTAT TTAATAACCA ATATCATACT TGGAGTGAGT TTAAGAAAGC	240
TATGTATCAA GAACGACAAG ATCAGTTTGA TAGATTGAAC AAAGTTACTT TTAATGATAC	300
AACACAGCCT TGGCAAACAT TTGCCAAGAA AACTACAAGC AGTGTAGATG AATTACAGAA	360

255

ATTAATGGAC GTTGCTGTTT GTAAGGATGC AGAACACAAT TACTACCATT GGAATAACTA	420
CAATCCAGAC ATAGATAGTG AAGTCCACAA GCTCAAGAGA GCAATCTTTA AAGCCTATCT	480
TGACCAAAACA AATGATTTTA GAAGTTCAAT TTTTGAGAAT AAAAAATAGT GTCTACTATT	540
AGGAAATAAA GTTTAAAAAG GTGATGAAGA ACAAACCAAG ATTCAAGCAG GAATTCCTAC	600
TGATAATGAA GTAAGTTATG ATCTTATTTA TCAGCAGGAA ACTCTTCCTG CAACAGGTTT	660
ATCAACTTCT GAGCTTACAG CTTTAGGCCT ATTAGCTGTT GGTAGTTTAG TTCTTTTGGT	720
TCATAATATG ACGGGAACAG TTTTGTGCTC CCTCTGAAA GTCATCATTT GATGGCTTTT	780
TTCTATATAG GGTAAAGAT AGGGTAAAAG GCTATCATCG GACAAAATAA AGAAGGCATG	840
ATATAATATA AAGTAGATTT CTATGTCATA AAACAAGAAC TGTTTGGACA TCATTCAATT	900
GAAAACCTCT TATGTTCAAA CAATAGTAAA ATAAATAGG GGATCTAAAT CCTTGCTATG	960
AAAGGAAAAA ACTCAATGGC TACTATTCAA TGGTTTCCTG GTCACATGTC TAAAGCTCGT	1020
CGACAGGTGC AGGAGAATTT AAAATTGTT GATTTTGTA CGATTTTAGT AGATGCACGC	1080
TTGCCTCTAT CTAGTCAAAA TCCTATGTTG ACCAAGATTG TTGGTGATAA ACCAAAACTC	1140
TTGATTTTAA ACAAGGCCGA CTTGGCTGAT CCAGCAATGA CCAAGGAATG GCGTCAGTAT	1200
TTTGAATCAC AAGGAATCCA GACGCTAGCT ATCAACTCCA AAGAGCAAGT GACTGTAAAA	1260
GTTGTAACAG ATGCGGCCAA GAAGCTCATG GCTGATAAGA TTGCTCGCCA GAAAGAACGT	1320
GGGATTCAGA TTGAAACCTT GCGTACTATG ATTATCGGGA TTCCAAACGC TGGTAAATCA	1380
ACTCTGATGA ACCGTTTGGC TGGTAAAAAG ATTGCTGTTG TTGGAAACAA GCCAGGGGTC	1440
ACAAAAGGTC AACAATGGCT TAAACCAAT AAAGACCTGG AAATCTTGA TACACCGGGG	1500
ATTCTCTGGC CTAAGTTTGA GGATGAACT GTTGCACTTA AGTTGGCATT GACTGGAGCT	1560
ATCAAAGACC AGTTGCTTCC TATGGATGAG GTTACCATT TTGGTATCAA TTATTTCAA	1620
GAACATTATC CAGAAAAGCT GGCTGAACGC TTCAAACAA TGAAAATTGA AGAAGAAGCG	1680
CCTGTGATTA TTATGGATAT GACCCGCGCC CTCGGTTTCC GTGATGACTA TGACCGTTT	1740
TACAGTCTCT TCGTGAAGGA AGTCCGTGAT GGCAACTCG GTAACATATAC CTTAGATACA	1800
TTGGAAGACC TCGATGGCAA CGATTAAAGA AATCAAAGAA TTCCTTGTGA CAGTCAAGGA	1860
GTTAGAAAAGC CCTATTTTTT TAGAGCTTGA AAAGGATAAT CGCTCAGGAG TTCAAAAGGA	1920
AATCAGCAAG CGTAAAAGAG CCATTCAAGC TGAATTAGAT GAAAATTGC GCTTGAATC	1980
CATGCTTTCT TATGAAAAAG AACTTTATAA GCAAGGATTG ACCTTAATTG CAGGTATTGA	2040
TGAGGTTGGT CGTGGTCCTC TTGCTGGTCC TGTAGTCGCT GCGGCCGTTA TTTTATCTAA	2100

256

AAATTGTAAG ATTAAGGTC TCAACGACAG CAAGAAAATT CCTAAAAAGA AACATCTGGA	2160
GATTTTCCAA GCCGTCAAG ACCAAGCCTT GTCGATTGGA ATTGGTATCA TAGATAATCA	2220
GGTCATCGAC CAAGTCAACA TCTATGAAGC AACCAACTA GCCATGCAAG AAGCAATCTC	2280
CCAGCTCAGC CCTCAACCAG AGCACCTTTT GATTGATGCC ATGAAACTGG ACTTGCCCAT	2340
TTCACAAACC TCCATTATCA AAGGAGATGC CAACTCCCTC TCTATCGCAG CAGCATCTAT	2400
AGTAGCCAAG GTAACACGTG ATGAATTGCT GAAAGAATAC GATCAGCAGT TCCCTGGCTA	2460
TGATTTTCGCT ACTAATGCAG GATATGGCAC AGCTAAACAT CTGGAAGGCC TCACAAAAC	2520
AGGAGTTACC CCAATTCACC GAACCAGCTT TGAACCCGTT AAATCACTGG TTTTAGGTAA	2580
AAAAGAAAGT TAATGAAAAG GAAATAACAT GGAGGAACAG TCGGAAATAG TCCGTTCTAA	2640
GAAAGAATTC GCCTTTGCAT CCAGACTAT ACTATCCCAA GTTGGTCGAG GAATCATTGT	2700
CGGCCATC GTTGAATTA TCGTCGGATC CTTTCGTTTC TTAATTGAAA AGGGCTTCCA	2760
CCTGATACAA GGAGTTTATC AAGATCAAGG GTACTTAGTG CGCAATCTTT TTGTACTGGT	2820
TTGTTTTAT ATACTCATCT GTTGGCTCAG TGCCAACTA ACACGGTCAG AAAAAGATAT	2880
TAAAGGCTCA GGAATTCCTC AAGTCGAAGC CGAACTGAAA GGCCTCATGT CCCTCAACTG	2940
GTGGGGCATT CTTTGAAAAA AATATGTGCT AGGTATTCTT GCTATTGCCA GTGGACTCAT	3000
GCTGGGTCGA GAGGGACCCA GCATTCAACT TGGAGCAGTT GGTGGTAAAG GAATTGCCAA	3060
GTGGCTCAA TCCAGTCCAG TAGAGGAACG TTCCTTGATT GCCAGTGGAG CTGCAGCAGG	3120
TTTAGCCGCA GCCTTTAATG CTCCTATTGC AGCACTCTC TTTGTTGTAG AAGAAGCTA	3180
TCACCATTTT TCGCGCTTTT TCTGGGTCTC AACTCTAGCA GCCAGCATCG TAGCAACTT	3240
TGTGTCTCTA CTCATGTTTC GTTTGACACC AGTATTGGAT ATGCCAGATA ACATTCTCTC	3300
CATGACCCTA GATCAGTATT GGATATATCT CGTCATGGGA ATTTTCCTTG GATTTTCAGG	3360
TTTTCTCTAT GAGAAAGCTG TATTAAACGT TGGAAGAGTT TATGACTTGA TTGGTCAAAA	3420
AATCCATTTG GATAGGGCTT ATTATCCCAT CTGGGCTTTT ATCCTTATCA TACCAGTCGG	3480
AATCTTCTTA CCTCAAATCA TTGGTGGCGG AAATCAGCTT GTCTTTTCTT TAACTGAACA	3540
AAATTTTAGT TTCCAAGTTT TATTAGCTTA CTTTTTAATC CGCTTTATTT GGAGTATGAT	3600
TAGCTATGGA AGTGGACTGC CAGGAGGAAT TTTCCTCCCC ATTTTAGCTC TTGGTTCTTT	3660
GCTTGGTGCC TTAGTTGGTG TTATCTGTGT CAATCTTGGA CTGTGCAGTC AAGAGCAATT	3720
CCCTATATTT GTCATTCTAG GAATGAGTGG CTATTTTGGA GCCATATCAA AAGCTCCCTT	3780
AACCGCTATG ATCCTCGTAA CTGAGATGGT AGGAGATATT CGCAACCTTA TGCCACTTGG	3840
TCTTGTCACT CTTGTTTCTT ATATTATCAT GGATTTGCTC AAAGGTACGC CAGTCTATGA	3900

257

AGCCATGCTG GAAAAAATGC TTCCAGAAGA AGTATCTAGC GAAGGAGAAG TTACACTTAT	3960
CGAAATACCA GTTCTGATA AAATGCTGG GAAACAAGTT CATGAACTCA ACTTACCACA	4020
CAACGTCCTC ATCACAATC AAGTCCATAA TGGCAAGAGC CAAACAGTTA ACGGCTCAAC	4080
CAGAATGTAT CTGGGTGATA TGATTCACCT GGTATTCCA AAAAGTGAAA TTGAAAAAGT	4140
CAAAGATTG TTGTTGTAGT ATGAGTATTT ACATAATTTA TGTTATGTAA ATGATCAGTT	4200
TGATTTATTT AGAAAACCGA TTCTCAGGAA TGAGATCGGT TATTTTTTAC TGATGAGGAA	4260
TTTTACATAT AAATAATTGA ACTTTATTAA AAATAAGACT ATAATTAAGT TAGAAATGAT	4320
AAAGTATAAA GCTAGAAAGG AGTTTACTGT ATCAAATCTG TACAGTAAGA TTAATAATCAT	4380
GAAAAAGAAA ACAATAGCAA TTATATAGAG AAATGAAATA GAAATAGGAT AAAACAATCA	4440
GGACAATCAA ATCAATTTCT AGCAATGTTT TAGAAGTCCA GATGTACTAT TCTAGTTTCA	4500
ATCTATTATA CAATGTGTTT TGTATCTCAT AGCTCCTTAT ATAGCTCTTC AGTTATGTAG	4560
TATTAACAGA AGTTTGTGG GTGAGATTTT TATTATTTTC CTTATCTGT TTGTTTGTA	4620
GGTCTAAGTC TTTTATCAC TTTGAAAAAC TCCTATAACA TCTTTCCGAA AAATAATAAT	4680
TTTCTTGAAA AATATACAAG TCTATGCTAT ACTACTAGTA TACTTACTTA TGGAGAAAAT	4740
ACATGAAACG TGAGATTTTA CTGGAACGAA TCGACAACT AAAACAATC ATGCCCTGGT	4800
AAGTTCTGGA ATACTACCAA	4820

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21338 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CTACGACATC ATGATTAAACA GTCATGCGCT ACTACCAACT GAGCTATGGC GGATAAAATA	60
GTCCGTACGG GATTCGAACC CGTGTACCG CCGTGAAAAG GCGGTGTCTT AACCCCTTGA	120
CCAACGGACC TTCTATCTGT AGCAGATATA ACCATTATAT CAATTTCTTG CTAATTGTCA	180
ATCACTTTTG AGATTTTTC TCTAAAATAT CTTTAAATTT TCTAATTTT AATCTTGAAA	240
TAGGACAACG ATGGTCTTCA TAGAAAACAA TTTCTAAGTT TTTTCGATCA ATTTCTCTGA	300
TATTACCTAT ATTTACCAA AATGACTTGT GAGGAGAATA AAATCGCTGA GTATGTTTGT	360
CCTTTCTCTG AATATCTGTC ATGGTACCAT AAAACTCTTT TGCAAAATTC TTACCAATAA	420

258

TGCGCAATTT ATGAGATACC CCTGTTGTTT CAATATACAA AATATCATGG TAAGGAATTT	480
TTAAATCATT TCCCTTGTA TGTAGTCGA AATAATCTAC AACATCTTCA TTTTCAAGTA	540
ACATACTCTT CGTGTAGAAG ATATTTTGCT CAATTCTCTT CTTAAACATC TCATCATTGA	600
TATCCTTATC AACAAAATCT AGGGCTGATA CCTGGTATTT ATAGGTTAGA GTCGCAAACT	660
CTGATCGACT AGTGATAAAG ACGATAATAG CGTAAGGATT GTAATGACGA ATGAGCTGAG	720
CCACTTCAAA TCCCTTTTTC TCAATTCCAT GAATATCGAT ATCTAGGAAA TAAAGCTGAT	780
TTACTTCATC ATTTTCAATG TATTCTTCAA ATTCACGGAC TTTTCCCGTT GTCTTGATG	840
ATATTGGAAT ATTCGATCTT TTCGAAATTT CATCCAATAT TCTCTCTAGT CTCACTTGAT	900
GTTCAATAAC ATCTTCTAAA ATTAAAACTT TCATTCAAAT TCCCTCTTAA ATCTAATGAT	960
TTGTCTAAAT GTACTGCCTT CCATCTCTGT TTCTAAAATA ATATTGTTGT ACTTATCTAG	1020
TAGTCTTTC ACATTATTTA ATCGACTCC GCGATTTCTT CCCTTAGTGG AGAATCCTAA	1080
GGCAAATAGA TCTCCTGAAG GAGTCATCGT CATTTTACAT GAATTCTGAA TCACAATAAC	1140
TGTTTCAGTT TCCATCTTAA TAACTGCTAC TTCCATCTGC TTTTATAGC TATCAGCCGA	1200
TCCTTCGACA GCATTATTCA ATAAAACGCT CATGATACGA ACCAAATCCA ATAGTTCAAT	1260
TGGAAGCTTG GTAATCGTAT CTTTACTTTC CAGTGTAAC TCTACCCAT TATTTGAGC	1320
ATAGACAATT GACTGAGCAA CCAAACCTCG TAAAGCTGAG TCTTCTATGT TGTCAAATC	1380
AAAGTAAGTG TACTTATCTG AACGCAATTT ATGATTTGCT TTGACTAAAA CTTCATTGTA	1440
AATTCTGTCA ATTTCTGTGA AATTACCACT GTCAATTGCC ATCTGCATGC TGACAAGCAT	1500
TCCAGCATAA TCATGTCGAA AACCACGGAT TTCATTATAC AGACCAACAA TTTCATCTGT	1560
GTAATTCTGT AAATGTTTCT GTTCAAATTT CTCTGCTTC AAAGCAATCT CTTTCTCCAT	1620
TTGAACCTTA TGAGAATTCA TTGCAAAGAA GGTCAAAGG AGAGAGATAA AGACAATAGA	1680
TGACAAAATA CTTCCAAAAC TATTCAAATG TTTAATCGTA CTTACCATAT CTGAAACGAA	1740
AGATACAATA TGTAGCAATA GTAAAGCAAA AAATACTTTT TTCAAGAAAG GATAAAGGTA	1800
GTCCTTGTC AATAGGCTA GTTCCAAATG GAAATAGTAA ATGATTTTTA ATGTAACAAA	1860
ATAGGTAAAC ACCGTCACAA CGAAAAAGAA TGGGAAATGA TATTGTAAAA CAAAATTATC	1920
TCCTGTTATA GAGGAGAAAA TTACGGACAG AAAGTTATGA GTGCTCTCAT ATAAAAGAGA	1980
TAGTAGTAAA CTTAGGAATA GTCCTCTATC CCTCTCATAC TGTTTCATCC ATCGAAAATA	2040
GGAATATAAG CCCAAAGGAA ATAAAAATCT TTCAATCCCT ATTTTATCTA AATATAGAAG	2100
ATAAAAGGAA AATTCAAGTA CTATTTCACT TAGTAATGTA TAAGCACCAA AAACGTATAA	2160
TTCTTTTCTA TTTATTCGAC CTTTACAAAT TAAACGGTAA CTGTGACTAA TAATTAAAAA	2220

259

ATGAACAATA ACTGTCCCAA ATCCAAGTAA ATCCATTACT CTTTCTCCTT ATTTCAATTAC	2280
TTTTTTTCGTA GGAAAAGAAA ATCAAGGATG ATTCTTGAAA TCCTCATCTC CCCACCTTTA	2340
ATCTTTTGTA AGTCTTTTTC CTTCAAAGCT ACAAACGTG CCAATTTAAC TGTGTTTTTC	2400
ATAATAAAAT CTCCTAAAT GTTTTTCTT GTAAGCTAAC TTACAAAAC CATTATACAA	2460
AATGGAATTT CGTTTAGAT AAAATCTCT CAACTGTCAT TTTTCTCTCC CAAAGTGATC	2520
TTTTTTAAGA AAAAAGCCGG GAAAATCCC AGCTTGCTA TTATATTGAT CCCAGCAGGA	2580
TTGGAACCTG CGACCGTTCG CTTAGAAGGC GAATGCTCTA TCCAGCTGAG CTATGAGACC	2640
TAATACAATT ATTCTACCAA AAATCAATT AAAAGTCAAT TTTCTATTTA TGGTAGGGGA	2700
ATCCCTGCTG AATCGTAAAA GCGCGATAGA TTTGTTCAAC AAGAACTAGT CTCATTAACT	2760
GATGGGGTAA GGTAGGCGA CCAAACTGA CAGAAAGATT GGCTCTATTT TTTACAGATG	2820
ATGATAATCC TAAACTTCCC CCAATAATAA AAGTAAGAGT AGAAAATCCT TTTATAGAAG	2880
TTTCTTCTAA CTGCTTACTA AATCTTCTG AGAAGAAAGT TTTCCCTTCA ATGGCTAACA	2940
CAATAACGAA ATCACGGTCA GCAATTTTTC ATAAAATCT CTGACCTTCT ATTTCTAAAA	3000
TCTTTTGAT TTTGATTCA CTGGCCTTAT CTGGTGT TTTCTCTGAT AACTCAATCA	3060
TTTCAAACCT AGCAAATCTA GAAATTCGTT TTGAATACTC TGCGATACCA TCTTTTAAAT	3120
ACTTTTCTTT CAGTTTCCCA ACTGTTACAA CTTTAATTTT CATGACTCTA TTCTAACATA	3180
TTCTCTATTT TTTACATCT TATTCACAAA ATAAAAATA GATTTCAATT AAGAAAATCA	3240
CAATTTCAA AGAGTTATCC ACAGTTTGTG TAAACTTTT GTGTTTAAAG TATAATTAAG	3300
CTAGTCAGTT TATACTTTCA GTAATTCAA CATATGGAGG CAAATATGAA ACATCTAAAA	3360
ACATTTTACA AAAAATGGTT TCAATTATTA GTCGTTATCG TCATTAGCTT TTTTAGTGGA	3420
GCCTTGGGTA GTTTTCAAT AACTCAACTA ACTCAAAAAA GTAGTGTAAC CAACTCTAAC	3480
AACAATAGTA CTATTACACA AACTGCCTAT AAGAACGAAA ATTCAACAAC ACAGGCTGTT	3540
AACAAAGTAA AAGATGCTGT TGTTCCTGTT ATTACTTATT CGGCAACAG ACAAAATAGC	3600
GTATTTGGCA ATGATGATAC TGACACAGAT TCTCAGCGAA TCTCTAGTGA AGGATCTGGA	3660
GTTATTTATA AAAAGAATGA TAAAGAAGCT TACATCGTCA CCAACAATCA CGTTATTAAT	3720
GGCGCCAGCA AAGTAGATAT TCGATTGTCA GATGGGACTA AAGTACCTGG AGAAATTGTC	3780
GGAGCTGACA CTTTCTCTGA TATTGCTGTC GTCAAAATCT CTTCAGAAAA AGTGACAACA	3840
GTAGCTGAGT TTGGTGATTC TAGTAAGTTA ACTGTAGGAG AAACTGCTAT TGCCATCGGT	3900
AGCCCGTTAG GTTCTGAATA TGCAAATACT GTCACCTAAG GTATCGTATC CAGTCTCAAT	3960

260

AGAAATGTAT CCTTAAATC GGAAGATGGA CAAGCTATTT CTACAAAAGC CATCCAAACT	4020
GATACTGCTA TTAACCCAGG TAACTCTGGC GGCCCACTGA TCAATATTCA AGGGCAGGTT	4080
ATCGGAATTA CCTCAAGTAA AATTGCTACA AATGGAGGAA CATCTGTAGA ACGTCTTGGT	4140
TTCGCAATTC CTGCAATGA TGCTATCAAT ATTATTGAAC AGTTAGAAAA AAACGAAAAA	4200
GTGACGCGTC CAGCTTTGGG AATCCAGATG GTTAATTTAT CTAATGTGAG TACAAGCGAC	4260
ATCAGAAGAC TCAATATTCC AAGTAATGTT ACATCTGGTG TAATTGTTTCG TTCGGTACAA	4320
AGTAATATGC CTGCCAATGG TCACCTTGAA AAATACGATG TAATTACAAA AGTAGATGAC	4380
AAAGAGATTG CTTTCATCAAC AGACTTACAA AGTGCTCTTT ACAACCATTTC TATCGGAGAC	4440
ACCATTAAGA TAACCTACTA TCGTAACGGG AAAGAAGAAA CTACCTCTAT CAAACTTAAC	4500
AAGAGTTCAG GTGATTAGA ATCTTAATTG ACATCTATGT AAAGAAAGCT TTACATAAGA	4560
GAAAGATGT GTTAGTGTAG AATCATGGAA AAATTTGAAA TGATTCTTAT CACAGATATA	4620
CAAAAAATC CCTATCAACC CCGAAAAGAA TTTGATAGAG AAAAAGTAGA TGAAGTAGCA	4680
CAGTCTATCA AAGAAATGG GGTCAATCAA CCGATTATTG TTCGTCAATC TCCTGTTATT	4740
GGTTATGAAA TCCTTGCAAG AGAGAGACGC TATCGGGCTT CACTTTTAGC TGGTCTACGG	4800
TCTATCCCAG CTGTTGTTAA ACAGATTTC AACCAGAGA TGATGGTCCA GTCCATTATT	4860
GAAATTTAC AGAGAGAAAA TTAAACCCA ATAGAAGAAG CACGCGCCTA TGAATCTCTC	4920
GTAGAGAAAG GATTCAACCA TGCTGAAATT GCAGATAAGA TGGGCAAGTC TCGTCCATAT	4980
ATCAGCAACT CCATTCGTTT ACTTCTCTTG CCAGAACAGA TTCTTTCAGA AGTAGAAAAT	5040
GGCAAACTAT CACAAGCCCA TGCGCGTTCC CTAGTTGGGT TAAATAAGGA ACAACAAGAC	5100
TATTTCTTTC AACGGATTAT AGAAGAAGAT ATTTCTGTAA GGAAATTAGA AGCTCTTCTG	5160
ACAGAGAAAA AACAAAAGAA ACAGCAAAAA ACTAATCATT TCATACAAAA TGAAGAAAAA	5220
CAGTTAAGAA AACTACTCGG ATTAGATGTA GAAATTAAAC TATCTAAAAA AGACAGTGGG	5280
AAAATCATT TTTCTTTTTC AAATCAAGAA GAATATAGTA GAATTATCAA CAGCCTGAAA	5340
TAAGGCTGTT CTTTATTTT TTTATCTCAC AAGGTTATCC ACTATGTTT TCGATAAAAA	5400
GCTTAATAAA TCAATAATT CTTCTTTTAT CCCCACCTG TGGATAAAGT TTGGTAACAT	5460
TGTGGATTAT TTTTCACAGC TTGTGGAAAA TTCTTGCTAT CTATGGTAAA ATATCTCTAG	5520
TATTAACTT TTAAATAGTA AAGGAGGAGA AAGGATTGAA AGAAAAACAA TTTTGGAATC	5580
GTATATTAGA ATTTGCACAA GAAAGACTGA CTCGATCCAT GTATGATTTC TATGCTATTC	5640
AAGCTGAAC CATCAAGGTA GAGGAAAATG TTGCCACTAT ATTTCTACCT CGCTCTGAAA	5700
TGGAAATGGT CTGGGAAAA CAACTAAAAG ATATTATTGT AGTAGCTGGT TTTGAAATTT	5760

261

ATGACGCTGA AATAACTCCC CACTATATTT TCACCAAACC TCAAGATACG ACTAGCTCAC	5820
AAGTTGAAGA AGCTACAAAT TTAACCTCTT ATAACATAG TCCAAAGTTA GTATCTATTC	5880
CTTATTCAGA TACGGGATTA AAAGAAAAGT ATACCTTTGA TAACTTTATT CAAGGGGATG	5940
GAAATGTTTG GGCTGTATCA GCCGCTTTAG CTGTCTCTGA AGATTTGGCT CTGACCTATA	6000
ACCCTCTTTT TATCTATGGA GGACCAGGCC TTGGTAAGAC TCACTTATTA AACGCTATTG	6060
GAAATGAAAT TCTAAAAAT ATTCCTAATG CGCGTGTTAA ATATATCCCT GCCGAAAGCT	6120
TTATTAATGA CTTTCTTGAT CACCTAAGAC TTGGGGAAT GGAAAAGTTT AAAAAGACCT	6180
ATCGTAGTCT TGATCTTTTG TTAATCGATG ATATCCAGTC ACTCAGCGGA AAAAAGTCG	6240
CAACTCAGGA AGAATTTTTC AATACCTTTA ACGCCCTTCA TGACAAGCAA AAACAGATTG	6300
TCCTAACGAG TGATCGTAGT CCAAACATC TAGAAGGGCT CGAGGAGAGG CTTGTCACGC	6360
GTTTTAGTTG GGGATTGACA CAAACTATCA CCCCCCTGA CTTTGAAACA CGTATTGCCA	6420
TTTTACAAAG TAAGACGGAA CATTTAGGCT ACAATTTCCA AAGTGATACT CTAGAATACC	6480
TAGCTGGGCA ATTTGATTCA AATGTTGAG ATCTTGAGGG AGCCATCAAC GACATCACTT	6540
TAATTGCCAG AGTAAAAAAT ATCAAGGATA TCACTATTGA TATTGCTGCA GAAGCCATTA	6600
GAGCCCGCAA ACAAGATGTT AGCCAAATGC TCGTCATCCC AATTGATAAA ATCCAAACTG	6660
AAGTTGGTAA CTTTATGCT GTTAGTATCA AAGAAATGAA GGGAAAGTAGA CGCCTTCAA	6720
ATATTGTTTT GGCCCGTCAA GTAGCCATGT ATTTATCTAG AGAACTAACA GATAATAGTC	6780
TTCCAAAAAT TGGGAAGGAA TTTGGGGGAA AAGATCATAC CACAGTCATT CATGCCCATG	6840
CCAAAATAAA ATCTTTGATT GATCAAGACG ATAATTTACG TTTAGAAATT GAATCAATCA	6900
AAAAGAAAAT CAAATAATTT GTGGATAACT TTTAGTTTTT TATCTTTTTT ATCCACATTT	6960
TTTAAACAAG CTAAAAAAT TGATATGACT TGTTTAAAGG CTGTTTTCCA CAGATTTCAC	7020
AGACTCTATT ATTACTATTA TCTTTCTAAT ACTAAAAATA AATAAAGGAG AATCCATGAT	7080
TCATTTTTCA ATTAATAAAA ATTTATTTCT ACAAGCATT AATACTACTA AGAGAGCTAT	7140
TAGTTCTAAA AATGCCATTC CTATTTTATC AACAGTAAAA ATTGACGTGA CCAATGAAGG	7200
TATTACTTTA ATTGTTTCAA ATGGTCAAAT TTCAATTGAA AATTTTATTT CTCAAAAAA	7260
TGAAGATGCT GGTGTTTAA TTACTTCTTT AGGTTTCGATC CTTCTTGAAG CTTCTTTCTT	7320
TATCAATGTA GTATCTAGTT TACCTGATGT AACTCTTGAT TTTAAAGAAA TTGAACAAAA	7380
TCAAATTGTT TTAACCAGTG GCAAATCAGA AATTACCCTA AAAGGAAAAG ATAGCGAACA	7440
ATATCCACGA ATCCAAGAAA TTTCAGCAAG CACTCCTTTA ATACTTGAAA CAAAATTACT	7500

262

CAAGAAAATT ATTAATGAAA CAGCCTTTGC TGCAAGTACA CAAGAGAGTC GTCCGATTTT	7560
AACAGGTGTC CACTTCGTAT TGAGTCAACA CAAAGAGTTA AAAACAGTTG CAACAGACTC	7620
TCATCGCCTA AGCCAGAAAA AATTGACTCT TGAAAAAAT AGTGATGATT TTGATGTCGT	7680
AATTCCTAGC CGTCTCTAC GCGAATTTTC AGCGGTATTT ACAGATGATA TCGAAACTGT	7740
AGAGATTTTC TTTGCCAATA ACCAAATCCT CTTTAGAAGC GAAAATATTA GCTTCTATAC	7800
TCGTCTCCTA GAAGGAAACT ATCCTGATAC AGATCGCTTG ATTCCAACAG ACTTTAACAC	7860
TACTATTACT TTTAATGTGG TAAACTTACG CCAGTCAATG GAGCGTGCCC GTCTTTTATC	7920
AAGTGCGACT CAAAATGGTA CTGTGAAACT TGAAATTAAG GATGGGGTTG TTAGCGCCCA	7980
TGTTCACTCT CCAGAAGTTG GTAAAGTAAA CGAAGAAATC GATACTGATC AGGTTACTGG	8040
TGAAGATTTG ACCATTAGTT TCAACCCAAC TTAAGTTGATT GATTCTCTTA AAGCTTTAAA	8100
TAGCGAAAAG GTGACTATTA GCTTTATCTC AGCTGTTCTG CCATTTACTC TTGTGCCAGC	8160
AGATACTGAC GAAGACTTCA TGCAGCTCAT TACACCAGTT CGTACAAATT AAGTGAAAGA	8220
GGTTGAGCCT GGCTCGCCTC TTTTATGATA TAATCGAAAA AGAAAAGGAG AGTAGTATGT	8280
ATCAAGTTGG AAATTTTGTG GAGATGAAAA AATCACACGC TTGTACAATC AAGTCGACTG	8340
GTA AAAAGGC TAATCGTTGG GAAATTACAC GTGTAGGAGC AGATATCAAA ATAAATGTA	8400
GTAATTGTGA GCATGTTGTC ATGATGGGGC GATATGATTT TGAGCGAAAA ATGAATAAAA	8460
TTATPGACTG AGAACCTTA GTTAGAGGGT TAGCACTTTA TCCCTTTTGG TGTATAATA	8520
TTAGGGATTG AAATGAAAAC GGAGAATGAG AAATATGGCT TTGACAGCAG GTATCGTTGG	8580
TTTGCCAAAC GTTGGTAAAT CAACACTATT TAATGCAATT ACAAAGCAG GAGCAGAGGC	8640
AGCAAACACTAC CCATTTGCCA CGATTGATCC AAATGTTGGA ATGGTGGAAG TTCCAGATCA	8700
ACGCCTACAA AACTAACTG AATGATAAC TCCTAAAAAG ACAGTTCCCA CAACATTTGA	8760
ATTTACAGAT ATTGCAGGGA TTGTAAAAGG AGCTTCAAAA GGAGAGGGGC TAGGGAATAA	8820
ATTCTTGGCC AATATTCGTG AAGTAGATGC GATTGTTTAC GTAGTTCGTG CTTTGTATGA	8880
TGAAAATGTA ATGCGCGAGC AAGGACGTGA AGACGCCTTT GTAGATCCAC TTGCAGATAT	8940
TGATACCATT AATCTGGAAT TGATTCTTGC TGACTTAGAA TCAGTGAACA AACGATATGC	9000
GCGTGTAGAA AAGATGGCAC GTACGCAAAA AGATAAGAA TCAGTAGCAG AATTCAATGT	9060
TCTTCAAAAAG ATTAAACCAG TCCTAGAAGA CGGAAATCA GCTCGTACCA TTGAATTTAC	9120
AGATGAGGAA CAAAAGGTTG TCAAAGGTCT TTTCTTTTGG ACGACTAAAC CAGTTCTTTA	9180
TGTAGCTAAT GTGGACGAGG ATGTGGTTTC AGAACCTGAC TCTATCGACT ATGTCAAACA	9240
AATTCGTGAA TTTGCAGCGA CAGAAAATGC TGAAGTAGTC GTTATTTCTG CGCGTGCTGA	9300

263

GGAAGAAATT TCTGAATTGA ATGATGAAGA TAAAAAAGAG TTTCTTGAAG CCATTGGTTT	9360
GACAGAATCA GGTGTAGATA AGTTGACGCG TGCAGCTTAC CACTTGCTTG GATTGGGAAC	9420
TTACTTCACA GCTGGTGAAA AAGAAGTTCG CGCTTGGACT TTCAAACGTG GTATGAAGGC	9480
TCCTCAAGCA GCTGGTATTA TCCACTCAGA CTTTGAAAAA GGCTTTATTC GTGCAGTAAC	9540
CATGTCATAT GAAGATCTAG TGAAATACGG ATCTGAAAAG GCCGTAAAAG AAGCTGGACG	9600
CTTGCGTGAA GAAGGAAAAG AATATATCGT TCAAGATGGC GATATCATGG AATTCCGCTT	9660
TAATGTCTAA AAATTAATAA ATGGTGTCAA TTAGGTTGGA AAAAAATTCC AACCTTTTG	9720
GCTTTTGAAA GGAAAAATAA ATGACCAAAT TACTTGTAGG CTGGGAAAT CCAGGGGATA	9780
AATATTTTGA AACAAAACAC AATGTTGGTT TTATGTTGAT TGATCAACTA GCGAAGAAAC	9840
AGAATGTCAC TTTTACACAC GATAAGATAT TTCAAGCTGA CCTAGCATCC TTTTTCCTAA	9900
ATGGAGAAAA AATTTATCTG GTTAAACCAA CGACCTTTAT GAATGAAAGT GGAAAAGCAG	9960
TTCATGCTTT ATTAACCTAC TATGGTTTGG ATATGACGA TTTACTTATC ATTTACGATG	10020
ATCTTGACAT GGAAGTTGGG AAAATTCGTT TAAGAGCAAA AGGCTCAGCA GGTGGTCATA	10080
ATGGTATCAA GTCTATTATT CAACATATAG GAACTCAGGT CTTTAACCGT GTTAAGATTG	10140
GAATTGGAAG ACCTAAAAAT GGTATGTCAG TTGTTTCATCA TGTTTTGAGT AAGTTTGACA	10200
GGGATGATTA TATCGGTATT TTACAGTCTG TTGACAAAGT TGACGATTCT GTAACTACT	10260
ATTTACAAGA GAAAAATTTT GAGAAAACAA TGCAGAGGTA TAACGGATAA ATGGTGACCT	10320
TATTAGATTT ATTCTCAGAA AATGATCAGA TTAaaaaATG GCATCAAAAT TTAACAGATA	10380
AGAAAAGACA ACTAATACTT GGTTTATCAA CATCTACTAA GGCTCTTGCA ATTGCAAGCA	10440
GTTTAGAAAA AGAAGATAGG ATTGTGTTAT TGACGTCAAC TTATGGAGAA GCAGAAGGAC	10500
TTGTTAGTGA TCTTATTTCT ATCTTGGGTG AGGAACTCGT CTATCCATTT TTGGTAGATG	10560
ATGCTCCTAT GGTGGAGTTT TTGATGTCTT CACAGGAAAA AATTATTTCA CGGGTTGAAG	10620
CCTTGCGTTT TTTGACTGAT TCATCTAAGA AAGGGATTTT AGTTTGTAAT ATCGCAGCAA	10680
GTGATTTGAT TTTACCGTCT CCCAATGCAT TCAAAGATAG TATTGTAAAA ATCTCAGTTG	10740
GTGAAGAATA TGATCAACAC GCGTTTATCC ATCAGTTAAA GGAAAATGGC TATCGAAAAG	10800
TTACTCAAGT ACAAACCTCAG GGCGAATTTA GTCTTCGAGG AGATATTTTA GATATTTTGG	10860
AAATATCCCA GTTAGAACCT TGTCGAATTG AGTMTTTTGG TGATGAAAT GATGGTATCA	10920
GGTCATTTGA AGTAGAAACA CAATTATCGA AAGAAAATAA GACAGAACTC ACTATCTTTC	10980
CAGCTAGTGA TATGCTTTTG AGAGAAAAGG ATTATCAACG AGGACAGTCA GCTTTAGAAA	11040

264

AACAAATTTC AAAAACTTTA TCACCTATTT TGAAATCATA CCTAGAAGAA ATTCTTTCAA	11100
GTTTTACCA AAAACAAAGT CATGCAGACT CTCGGAAGTT TTTATCTTTG TGCTATGATA	11160
AGACATGGAC TGTCTTTGAT TATATGAAA AAGATACTCC AATATCTTTT GATGATTATC	11220
AAAAATTGAT GAATCAGTAT GAAGTCTTTG AAAGAGACTT AGCGCAGTAC TTTACAGAAG	11280
AATTACAGAA TAGTAAAGCA TTTTCTGATA TGCAGTATTT TTCTGATATT GAACAAATCT	11340
ATAAAAAACA AAGTCCAGTG ACCTTTTCTCT CTAATCTTCA AAAGGGTTTA GGAAATCTCA	11400
AATTGACAA AATTTATCAA TTCAATCAAT ATCCTATGCA GGAATTTTTC AATCAGTTTT	11460
CTTTTCTAAA AGAAGAAATT GAACGATATA AAAAAATGGA TTACACCATT ATTCTGCAGT	11520
CTAGCAATTC AATGGGAAGT AAAACATTGG AGGATATGTT AGAGGAATAT CAGATTAAAT	11580
TGGATTCTAG AGATAAGACA AATATCTGTA AAGAATCTGT AAACCTAATA GAGGGTAATC	11640
TCAGACATGG TTTTCATTTT GTAGATGAAA AGATTTTATT GATAACTGAA CATGAGATTT	11700
TTCAAAGAA ATTAAAGCGT CGTTTTCGAA GACAACATGT TTCAAATGCA GAGAGATTAA	11760
AAGATTACAA TGAACCTGAA AAAGGGGACT ATGTTGTCCA TCATATCCAT GGGATTGGTC	11820
AATATCTAGG AATTGAAACC ATTGAAATCA AGGGAATTCA TCGCGATTAT GTCAGTGTCC	11880
AATACCAAAA TGGTGATCAA ATTTCTATCC CCGTGAACA GATTCATCTA CTGTCCAAAT	11940
ATATTTCAAG TGATGGTAAA GCTCCAAAC TCAATAAATT AAATGACGGT CATTTTAAAA	12000
AGGCCAAGCA AAAGGTTAAG AACCAGGTAG AGGATATAGC TGATGATTTA ATCAAATCT	12060
ACTCTGAACG TAGTCAGTTG AAGGGTTTGT CTTTCTCAGC TGATGATGAT GATCAAGATC	12120
CCTTTGATGA TGCTTTCCCT TATGTTGAAA CGGATGATCA ACTTCGTAGT ATTGAGGAAA	12180
TCAAGAGGGA TATGCAGGCT TCTCAGCCAA TGGATCGACT TTTAGTTGGG GATGTTGGTT	12240
TTGGAAGAC TGAAGTTGCT ATGCGTGCAG CCTTTAAAGC AGTCAATGAT CACAAACAGG	12300
TTGTCATTCT AGTTCCGACG ACGGTTTTAG CGCAACAGCA CTATACGAAT TTTAAGGAAC	12360
GATTCACAAA TTTTGCAGTT AATATTGATG TGTGAGTCG CTTTAGAAGT AAAAAAGAGC	12420
AGACTGCAAC ACTTGAAAAA TTGAAAAACG GTCAAGTCGA TATTTTGATT GGAACACATC	12480
GTGTTTTGTC AAAAGATGTT GTGTTGCTG ATTTGGGCTT GATGATTATT GATGAGGAAC	12540
AGCGATTTGG TGTCAAGCAT AAGGAACTT TGAAAGAACT GAAGAAACAA GTGGATGTCC	12600
TAACCTTGAC CGCTACGCCA ATCCCTCGTA CCCTCCATAT GTCTATGCTG GGAATCAGAG	12660
ATTTATCTGT TATTGAACT CCGCCGACTA ATCGCTATCC GTTCAGACC TATGTTTTGG	12720
AAAAGAATGA TAGTGTCATT CGTGATGCTG TCTTGCGTGA AATGGAGCGT GGAGGTCAAG	12780
TTTATTATCT TTACAACAAA GTTGACACAA TTGTTTCAGAA GGTTTCAGAA TTACAGGAGT	12840

265

TGATTCGGGA GGCTTCGATT GGATATGTTT ATGGTCGAAT GAGTGAAGTC CAGTTGGAAA	12900
ATACTCTATT AGACTTTATT GAGGGACAAT ACGATATCTT GGTGACGACT ACTATTATTG	12960
AGACAGGGGT GGACATPCCA AATGCTAATA CTTTATTTAT TGAAAATGCG GACCATATGG	13020
GCTTGTC AAC CTTATATCAG TTAAGAGGAA GAGTCGGTCG TAGTAATCGT ATTGCTTATG	13080
CTTATCTCAT GTATCGTCCA GAAAATCAA TCAGTGAAGT CTCTGAAAAG AGATTAGAAG	13140
CGATTAAAGG ATTTACAGAA TTGGGCTCTG GCTTTAAGAT TGCAATGCCA GATCTTTCGA	13200
TTCTGCGAGC AGGAAATCTT TTAGGAAAAT CCCAGTCTGG TTTCATTGAT TCTGTTGGTT	13260
TTGAATTGTA TTCGCAGTTA TTAGAGGAAG CTATTGCTAA ACGAAACGGT AATGCTAACG	13320
CTAACACAAG AACCAAAGGG AATGCTGAGT TGATTTTGCA AATTGATGCC TATCTTCCTG	13380
ATACTTATAT TTCTGATCAA CGACATAAGA TTGAAATTTA CAAGAAAATT CGTCAAATTG	13440
ACAACCGTGT CAATTATGAA GAGTTACAAG AGGAGTTGAT AGACCGTTTT GGAGAATACC	13500
CAGATGTAGT AGCCTATCTG TTAGAGATTG GPTTGGTCAA ATCATACTTG GACAAGTCT	13560
TTGTTCACG TGTGGAAAGA AAAGATAATA AAATTACAAT TCAATTTGAA AAAGTCACTC	13620
AACGACTGTT TTTAGCTCAA GATTATTTTA AAGCTTTATC CGTAACGAAC TTAAAAGCAG	13680
GCATCGCTGA GAATAAGGGA TTAATGGAGC TTGTATTTGA TGTCCTAAAAT AAGAAAGATT	13740
ATGAAATTTT AGAAGGTTTG CTGATTTTGG GAGAAAGTTT ATTAGAGATA AAAGAGCTTA	13800
AGGAAGAAAA TTCCATTGTA TATTTTCTT CTATAAAATA GATAAAATG GTACAATAAT	13860
AAATTGAGGT AATAAGGATG AGATTAGATA AATATTTAAA AGTATCGCGA ATTATCAAGC	13920
GTCGTACAGT CGCAAAGGAA GTAGCAGATA AAGGTAGAAT CAAGGTTAAT GGAATCTTGG	13980
CCAAAAGTTC AACGGACTTG AAAGTTAATG ACCAAGTTGA AATTCGCTTT GGCAATAAGT	14040
TGCTGCTTGT AAAAGTACTA GAGATGAAAG ATAGTACAAA AAAAGAAGAT GCAGCAGGAA	14100
TGTATGAAAT TATCAGTGAA ACACGGGTAG AAGAAAATGT CTAAAAATAT TGTACAATTG	14160
AATAATTCTT TTATTCAAAA TGAATACCAA CGTCGTCGCT ACCTGATGAA AGAACGACAA	14220
AAACGGAATC GTTTTATGGG AGGGGTATTG ATTTTGATTA TGCTATTATT TATCTTGCCA	14280
ACTTTTAATT TAGCGCAGAG TTATCAGCAA TTAATCCAAA GACGTCAGCA ATTAGCAGAC	14340
TTGCAAACTC AGTATCAAAC TTTGAGTGAT GAAAAGGATA AGGAGACAGC ATTTGCTACC	14400
AAGTTGAAAG ATGAAGATTA TGCTGCTAAA TATACACGAG CGAAGTACTA TTATTCTAAG	14460
TCGAGGGAAA AAGTTTATAC GATTCCTGAC TTGCTTCAAA GGTGATAAAA TGGAAAATTT	14520
ATTAGACGTA ATAGAGCAAT TTTTGAGTTT GTCAGATGAA AAGCTGGAAG AATTGGCTGA	14580

265

TAAAAATCAA TTATTGCGTT TACAAGAAGA AAAGGAAAGG AAGAATGCGT AAATTCCTTAA	14640
TTATTTTGTT GCTACCAAGT TTTTGGACCA TTCAAAAGT CGTTAGCACA GAAAAAGAAG	14700
TCGTCTATAC TTCGAAAGAA ATTTATTACC TTTCACAATC TGACTTTGGT ATTTATTTTA	14760
GAGAAAAATT AAGTTCTCCC ATGGTTTATG GAGAGGTTCC TGTATTATGCG AATGAAGATT	14820
TAGTAGTGGA ATCTGGGAAA TTGACTCCCA AAACAAGTTT TCAAATAACC GAGTGGCGCT	14880
TAAATAAACA AGGAATTCCA GTATTTAAGC TATCAAATCA TCAATTTATA GCTGCGGACA	14940
AACGATTTTT ATATGATCAA TCAGAGGTAA CTCCAACAAT AAAAAAAGTA TGGTTAGAAT	15000
CTGACTTTAA ACTGTACAAT AGTCCTTATG ATTTAAAAGA AGTGAAATCA TCCTTATCAG	15060
CTTATTCGCA AGTATCAATC GACAAGACCA TGTTTGTAGA AGGAAGAGAA TTTCTACATA	15120
TTGATCAGGC TGGATGGGTA GCTAAAGAAT CAACTTCTGA AGAAGATAAT CGGATGAGTA	15180
AAGTTCAAGA AATGTTATCT GAAAAATATC AGAAAGATTC TTTCTCTATT TATGTTAAGC	15240
AACTGACTAC TGGAAAAGAA GCTGGTATCA ATCAAGATGA AAAGATGTAT GCAGCCAGCG	15300
TTTTGAAACT CTCTTATCTC TATTATACGC AAGAAAAAT AAATGAGGGT CTTTATCAGT	15360
TAGATACGAC TGTAAAATAC GTATCTGCAG TCAATGATTT TCCAGGTTCT TATAAACCAG	15420
AGGGAAGTGG TAGTCTTCCT AAAAAAGAAG ATAATAAGA ATATTCTTTA AAGGATTTAA	15480
TTACGAAAGT ATCAAAAGAA TCTGATAATG TAGCTCATAA TCTATTGGGA TATTACATTT	15540
CAAACCAATC TGATGCCACA TTCAAATCCA AGATGTCTGC CATTATGGGA GATGATGGG	15600
ATCCAAAAGA AAAATTGATT TCTTCTAAGA TGGCCGGGAA GTTTATGGAA GCTATTTATA	15660
ATCAAAATGG ATTTGTGCTA GAGTCTTTGA CTAAAACAGA TTTTGATAGT CAGCGAATTG	15720
CCAAAGGTGT TTCTGTTAAA GTAGCTCATA AAATTGGAGA TGCCGATGAA TTTAAGCATG	15780
ATACGGGTGT TGTCTATGCA GATTCTCCAT TTATTCTTTC TATTTTCACT AAGAATTCTG	15840
ATTATGATAC GATTTCTAAG ATAGCCAAGG ATGTTTATGA GGTTCTAAAA TGAGGGAAACC	15900
AGATTTTTTA AATCATTTTC TCAAGAAGGG ATATTTCAA AAGCATGCTA AGGCGGTCT	15960
AGCTCTTTCT GGTGGATTAG ATTCCATGTT TCTATTTAAG GTATTGTCTA CTTATCAAAA	16020
AGAGTTAGAG ATTGAATTGA TTCTAGCTCA TGTGAATCAT AAGCAGAGAA TTGAATCAGA	16080
TTGGGAAGAA AAGGAATTAA GGAAGTTGGC TGCTGAAGCA GAGCTTCCTA TTTATATCAG	16140
CAATTTTTCA GGAGAATTTT CAGAAGCGCG TGCACGAAAT TTTCGTTATG ATTTTTTTCA	16200
AGAGGTCATG AAAAAGACAG GTGCGACAGC TTTAGTCACT GCCCACCATG CTGATGATCA	16260
GGTGGAAACG ATTTTATGCG GCTTGATTCG AGGAACTCGC TTGCGCTATC TATCAGGAAT	16320
TAAGGAGAAG CAAGTAGTCG GAGAGATAGA AATCATTCGT CCCTTCTTGC ATTTTCAGAA	16380

267

AAAAGACTTT CCATCAATTT TTCACTTTGA AGATACATCA AATCAGGAGA ATCATTATTT	16440
TCGAAATCGT ATTCGAAATT CTTACTTACC AGAATTGGAA AAAGAAAATC CTCGATTTAG	16500
GGATGCAATC TTAGGCATTG GCAATGAAAT TTTAGATTAT GATTTGGCAA TAGCTGAATT	16560
ATCTAACAAT ATTAATGTGG AAGATTTACA GCAGTTATTT TCTTACTCTG AGTCTACACA	16620
AAGAGTTTTA CTTCAAACCT ATCTGAATCG TTTTCCAGAT TTGAATCTTA CAAAAGCTCA	16680
GTTTGCTGAA GTTCAGCAGA TTTTAAAATC TAAAAGCCAG TATCGTCATC CGATTAAAAA	16740
TGGCTATGAA TTGATAAAAG AGTACCAACA GTTTCAGATT TGTAAAATCA GTCCGCAGgC	16800
TGATGAAAAG GAAGATGAAC TTGTGTTACA CTATCAAAT CAGGTAGCTT ATCAAGGATA	16860
TTTATTTTCT TTTGGACTTC CATTAGAAGG TGAATTAATT CAACAAATAC CTGTTTCACG	16920
TGAAACATCC ATACACATTC GTCATCGAAA AACAGGAGAT GTTTTGATTA AAAATGGGCA	16980
TAGAAAAAAA CTCAGACGTT TATTTATTGA TTTGAAAATC CCTATGGAAA AGAGAACTC	17040
TGCTCTTATT ATTGAGCAAT TTGGTGAAT TGTCTCAATT TTGGGAATTG CGACCAATAA	17100
TTTGAGTAAA AAAACGAAAA ATGATATAAT GAACACTGTA CTTTATATAG AAAAAATAGA	17160
TAGGTAAAAA ATGTTAGAAA ACGATATTAA AAAAGTCCTC GTTTCACACG ATGAAATTAC	17220
AGAAGCAGCT AAAAACTAG GTGCTCAATT AACTAAAGAC TATGCAGGAA AAAATCCAAT	17280
CTTAGTTGGG ATTTTAAAG GATCTATTCC TTTTATGGCT GAATTGGTCA AACATATTGA	17340
TACACATATT GAAATGGACT TCATGATGGT TTCTAGCTAC CATGGTGGAA CAGCAAGTAG	17400
TGGTGTATC AATATTAAAC AAGATGTGAC TCAAGATATC AAAGGAAGAC ATGTTCTATT	17460
TGTAGAAGAT ATCATTGATA CAGGTCAAAC TTTGAAGAAT TTGCGAGATA TGTTTAAAGA	17520
AAGAGAAGCA GCTTCTGTTA AAATTGCAAC CTGTTGGAT AAACCAGAAG GACGTGTTGT	17580
AGAAATTGAG GCAGACTATA CTTGCTTTAC TATCCCAAAT GAGTTTGTAG TAGGTTATGG	17640
TTTAGACTAC AAAGAAAATT ATCGTAATCT TCCTTATATT GGAGTATTGA AAGAGGAAGT	17700
GTATTCAAAT TAGAAAGAAT AATCTTTAAT GAAAAACAA AATAATGGTT TAATTAAAAA	17760
TCCTTTTCTA TGGTTATTAT TTATCTTTT CCTTGTGACA GGATTCCAGT ATTTCTATTC	17820
TGGGAATAAC TCAGGAGGAA GTCAGCAAAT CAACTATACT GAGTTGGTAC AAGAAATTAC	17880
CGATGGTAAT GTAAAAGAAT TAACTTACCA ACCAAATGGT AGTGTATATCG AAGTTTCTGG	17940
TGTCTATAAA AATCCTAAAA CAAGTAAAGA AGAAACAGGT ATTCAGTTTT TCACGCCATC	18000
TGTTACTAAG GTAGAGAAAT TTACCAGCAC TATTCTTCCT GCAGATACTA CCGTATCAGA	18060
ATTGCAAAAA CTTGCTACTG ACCATAAAGC AGAAGTAACT GTTAAGCATG AAAGTTCAAG	18120

268

TGGTATATGG ATTAATCTAC TCGTATCCAT TGTGCCATTT GGAATTCTAT TCTTCTTCCT	18180
ATTCTCTATG ATGGGAAATA TGGGAGGAGG CAATGGCCGT AATCCAATGA GTTTTGGACG	18240
TAGTAAGGCT AAAGCAGCAA ATAAAGAAGA TATTAAAGTA AGATTTTCAG ATGTTGCTGG	18300
AGCTGAGGAA GAAAAACAAG AACTAGTTGA AGTTGTTGAG TTCTTAAAAG ATCCAAAACG	18360
ATTCACAAAA CTGGAGCCC GTATTCCAGC AGGTGTTCTT TTGGAGGGAC CTCGGGGGAC	18420
AGGTAAACT TTGCTTGCTA AGGCAGTCGC TGGAGAAGCA GGTGTTCCAT TCTTTAGTAT	18480
CTCAGGTTCT GACTTTGTAG AAATGTTTGT CGGAGTTGGA GCTAGTCGTG TTCGCTCTCT	18540
TTTTGAGGAT GCCAAAAAAG CAGCACCAGC TATCATCTTT ATCGATGAAA TTGATGCTGT	18600
TGGACGTCAA CGTGGAGTCG GTCTCGCGG AGGTAATGAC GAACGTGAAC AAACCTTGAA	18660
CCAACCTTTG ATTGAGATGG ATGGTTTGA GGGAAATGAA GGGATTATCG TCATCGCTGC	18720
GACAAACCGT TCAGATGTAC TTGACCCTGC CCTTTTGCCT CCAGGACGTT TTGATAGAAA	18780
AGTATTGGTT GGTGCTCCTG ATGTTAAAG TCGTGAAGCA ATCTTGAAAG TTCACGCTAA	18840
GAATAAGCCT TTAGCAGAAG ATGTTGATTT GAAATTAGTG GCTCAACAAA CTCCAGGCTT	18900
TGTTGGTGCT GATTTAGAGA ATGTCTTGAA TGAAGCAGCT TTAGTTGCTG CTCGTCGCAA	18960
TAAATCGATA ATTGATGCTT CAGATATTGA TGAAGCAGAA GATAGAGTTA TTGCTGGACC	19020
TTCTAAGAAA GATAAGACAG TTTCACAAAA AGAACGAGAA TTGGTTGCTT ACCATGAGGC	19080
AGGACATACC ATGTTGGTC TAGTCTTGTC GAATGCTCGC GTTGTCCATA AGGTTACAAT	19140
TGTACCACGC GGCCGTGCAG GCGGATACAT GATTGCACTT CCTAAAGAGG ATCAAATGCT	19200
TCTATCTAAA GAAGATATGA AAGAGCAATT GGCTGGCTTA ATGGGTGGAC GTGTAGCTGA	19260
AGAAATTATC TTTAATGTCC AAACCACAGG AGCTTCAAAC GACTTTGAAC AAGCGACACA	19320
AATGGCACGT GCAATGGTTA CAGAGTACGG TATGAGTGAA AAACCTGGCC CAGTACAATA	19380
TGAAGGAAAC CATGCTATGC TTGGTGACACA GAGTCCTCAA AAATCAATTT CAGAACAAAC	19440
AGCTPATGAA ATTGATGAAG AGGTTGCTTC ATTATTAAAT GAGGCACGAA ATAAAGCTGC	19500
TGAAATTATT CAGTCAAATC GTGAAACTCA CAAGTTAATT GCAGAAGCAT TATTGAAATA	19560
CGAAACATTG GATAGTACAC AAATTAAAGC TCTTTACGAA ACAGGAAAGA TGCCTGAAGC	19620
AGTAGAAGAG GAATCTCATG CACTATCCTA TGATGAAGTA AAGTCAAAAA TGAATGACGA	19680
AAAATAACCC TGAGAGAGGC TGGAGCCTCT CTTTTTGTG CAGTTTAGGA GCTAAAGGGA	19740
ACAGAATGGA GAAAAATGAA CAAATGTGTT TTCTAATCTG TTAGACTGTA TCTAGAAAGG	19800
GGAAATTTAT GATTAAAGAA TTGTATGAAG AAGTCCAAGG GACTGTGTAT AAGGTAGAA	19860
ATGAATATTA CCTTCATTTA TGGGAATTGT CGGATTGGGA GCAAGAAGGC ATGCTCTGCT	19920

269

TACATGAATT GATTAGTAGA GAAGAAGGAC TGGTAGACGA TATTCACGT TTAAGGAAAT	19980
ATTTCAAGAC CAAGTTTCGA AATCGAATT TAGACTATAT CCGTAAACAG GAAAGTCAGA	20040
AGCGTAGATA CGATAAAGAA CCCTATGAAG AAGTGGGTGA GATCAGTCAT CGTATAAGTG	20100
AGGGGGGTCT CTGGCTAGAT GATTATTATC TCTTTCATGA AACACTAAGA GATTATAGAA	20160
ACAAACAAAG TAAAGAGAAA CAAGAAGAAC TAGAACGCGT CTTAAGCAAT GAACGATTTC	20220
GAGGGCGTCA AAGAGTATTA AGAGACTTAC GCATTGTGTT TAAGGAGTTT ACTATCCGTA	20280
CCCACTAGTA ACTCATGCAA AAAAAATGAA AAAAATTAGA AAAAGTAGTT GACAAAGTTT	20340
GAAAAGGCTG TATAATAGTA AGAGTTGAAA ATAACAAC TCAGTCCGTTG GTCAAGGGGT	20400
TAAGACACCG CCTTTTCACG GCGGTAACAC GGGTTCGAAT CCCGTACGGA CTATGGTATG	20460
TTGCGTCAGG ACCACTTGAT GAAAAAAGT TTAACAAAAC TTAACAACTC TCAAAAAAGT	20520
GTGACAAGC GAAAGCAGTT GTGATATACT AATATAGTTG TCGCTTGAGA GAAGCAAGTG	20580
ACAAAGACCT TTGAAACTG AACAGACGA ACCAATGTGC AGGGCGCTAC AACGTAAGTT	20640
GTAGTACTGA ACAATGAAAA AAACAATAAA TCTGTCAGTG ACAGAAATGA GTAAGAACTC	20700
AAACTTTTTA ATGAGAGTTT GATCCTGGCT CAGGACGAAC GCTGGCGGCG TGCCTAATAC	20760
ATGCAAGTAG AACGCTGAAG GAGGAGCTTG CTTCTCTGGA TGAGTTGCGA ACGGGTGAGT	20820
AACGCGTAGG TAACCTGCCT GGTAGCGGGG GATAACTATT GGAAACGATA GCTAATACCG	20880
CATAAGAGTA GATGTTGCAT GACATTTGCT TAAAGGTGC ACTTGCACTA CTACCAGATG	20940
GACCTGCGTT GTATTAGCTA GTTGGTGGG TAACGGCTCA CCAAGGCGAC GATACATAGC	21000
CGACCTGAGA GGGTGATCGG CCACACTGGG ACTGAGACAC GGCCAGACT CCTACGGGAG	21060
GCAGCAGTAG GGAATCTTCG GCAATGGACG GAAGTCTGAC CGAGCAACGC CGCGTGACTG	21120
AAGAAGGTTT TCGGATCGTA AAGCTCTGTT GTAAGAGAAG AACGAGTGTG AGAGTGGA	21180
GTTCACTG TGACGGTATC TTACCAGAAA GGGACGGCTA ACTACGTGCC AGCAGCCGCG	21240
GTAATACGTA GGTCCCGAGC GTTGTCCGGA TTTATTGGGC GTAAAGCGAG CGCAGGCGGT	21300
TAGATAAGTC TGAAGTAAA GCCTGTGGCT TAACCATA	21338

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6273 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

270

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TGTTTTTAAA GAGCCGTGTC TGGATAGACT TTCGGACGCA ACGCTCTATT AGATAATGAA	60
CTGCCTATAC ACAAGATTTC TAACCTTAGT CGACATGAGC TGAAACCTCT TATTTGTTAA	120
GTAGTTCACA AAATATTATA CACCTATTTT ATGAATAGTC AACTGTCTTT ACAGTAAAAT	180
TTTAGAAAAT CATGAAAATT TTCTCTTTCT TTCCATTTTA AGTGACATTC AGTCATTCTC	240
ACATCAAAAA AGCCCAGACG AAATTGTCTG AGCATTCCTT TATCTAGTCG TTTAAGGAAG	300
TTGAGTTCAG TATGTTTAAA GTCTCTGTCC CATCATTTCT TCAACAAACC TTGTTCTTGG	360
AGAAACTCCT TGGCTACTTG CTTTGCTGAC TTGCCTTCAA CACCGACTTG GTAGTTGAGC	420
TGGCTCATCT GGCTTTCTGT AATCTTACCA GCCAATGTAT TAAGAACTCT TTCCAACCTCT	480
GGGTGTTTCT TGAGAAGAGC TTCTTTTCATG AGTGGAGCCC CTTGATAAGG TGGGAAGAGT	540
TGCTTGTCAT CTTCCAAGAC CTGTAAATCA TAACGCTCCA ATTCCGCATC AGTCGAATAG	600
GCATCCGTGA TTTGAATATC CCCTGACTGA ATAGCCTGAT AGCGAAGGGC TGGCTCAATG	660
GTCGTACAT TGAGATTGAG ACCATACATT GATTGCAAGC CCTTATTTCC ATCTTCACGG	720
TCGTTAAACT CGAGTGTAAG ACCTGCCTTC AACTGCCCTT CCACTTTTTT CAAGTCTGAA	780
ATGGTCTTCA AGCCATATTC TTGAGCAATC TTTTTCGGAA CAGCTACAGC ATAGGTGTTT	840
TGATAAGACA TGGGTTTGAG ATAGGCTAGA TGATCCTGCT TAGCAATGCC ATCACGCGCC	900
ACCTGATAAA CCTGTTCTGG TTCATGACTC ACCTTGGGTG ATGGTTGAAG CAAACTTTCA	960
GTCACCGTAC CAGTAAATTC AGGATAGATG TCAATATCGC CTTTTTTCAG AGCTTCATAA	1020
AGGAAGCTTG TCTTCCCAA ATTTCGGTTA ACAGTCGCAG TCATGCTGGT ATTTTCTTCA	1080
ATCAGCAACT TATACATATT GGCCAAAATT TCTGGTTCTG GACCTATTTT CCCAGCAATA	1140
ACCAAGTTTT CCTTCTCTTT TTGAACCAA AGAGCTGGAC TATAAGACAG ACCCAGTAAT	1200
AAAGCCACCA AGGCAAAACC TGAGAAAATC GTCCGTAATT TTGCTTTTTC CATCACTTTT	1260
AGTAGGAAGT TAAAGGCAAT GGCTAGCACT GCAGAAGAAA GTGCCCCAAT CAAAATCAAA	1320
CTGGCATTAT TACGGTCAAT TCCCAAAGA ATAAAGGAAC CTAGTCCCCC TGCACCAATC	1380
AAGCCGCCA AGGTTGCCGT ACCGATAATC AAAACAGCTG CCGTCCGAAT CCCAGACATG	1440
ATAACAGGCA TGGCGAGTGG AATTTCAAAT TTCTTGAGAC GTTCCCATCT GGTCAATCCA	1500
AAGGCAATCC CAGCCTCTTG CAGGTTTCGA TCAATTCCCT TCAGCCCAGT GATAGTATTT	1560
TGCAAAATAG GGAAAATCGC ATAAATCACT AGAGCTGTCA AAGCCGGCAA GGTCCCAATT	1620
CCCATCAAAG GGATAAGAG CCCCAACAAG GCCAGAGACG GGATGGTCTG GAAAATACCT	1680
GCAATCTGCA AGACCCAGTC GGCCAGCTTC TCATGATAGC GAAGAAAAAC AGCCAAGGGA	1740

271

ATCGCAAGCA AAATAGCTAG TAACAAGGTC AAAAGCGACA ACTGCAAATG TTGAGATAGA	1800
GCTGTCAACC AATCACTAAA ACGATCCTGA AAAGTTGCAA TTAAATTAGT CATGAACACT	1860
ACCTCCAAAC AAGTCTGCTA CAAAGTCTGT TGCAGGCGCT TTTAAAATTG TCTCGGGATT	1920
CGCTACCTGG CGAATTTCTC CATCCTGCAA GACAGCAATA CGGTCCGCCA ACTTCAAGGC	1980
TTTCATCCGTA TCATGGGTTA CAAAAATCGT TGTCATCCCA AACTCTTTAT GCAATTCTTT	2040
TGTCAGAACCC TGCAACTGTT TTCTCGAAAT AGCATCCAAG GCCGAAAAGG GTTCATCCAT	2100
GAGGAAAATC TTGGGCTGAC CAATCATAGC TCGGACAATA CCGACCCGTT GCTGTTCTCC	2160
ACCAGATAAT TCACTAGGTA AGCGATGCCC ATACTCGGCT ACTGGTAAAC CAACCTTAGC	2220
CAAAGCTCT TCTGTTTTCT TCGTAATTC TTCTTGCTC CACCCCTTCA TTTCAGGAAT	2280
GAGAGCAATA TTTCCGCAA CTGTTAGATT TGGAAAAAGA GCAATAGCCT GTAAAACATA	2340
ACCAGTAGAA AGACGAAGTT CACGCTCATC ATAGTCTTG ATGCGCTTCC CATCCATATA	2400
AATATTTCCA TCAGTTGGTT CCAAAGACG GTTAATCATC TTGAGCATGG TCGTCTTACC	2460
TGACCCAGAA GGCCCTACTA AAACCATAAA TTCCCATCC TCAATCTGTA AGTTGACATC	2520
TCTCAAGACA TCCTTTCTG TGTAGCGCAG TGCTACATTT TTGTATTCAA TCATTTCTTG	2580
TCCTCAATTT AAAACTTCCC TCGATTGGTC AAGTCTTCTA CCTTAGGCAT AACTTCCTTA	2640
TTATCCCAAT GCTCCACAAT TTTCCGTTTCTTAAACGGA AGATATCGTA CTGGGCATAA	2700
GCAACGCCAT CAATCTGAGT CTGACCATAG CTAACCACAT AGTTTCCTTG TCCTAAGAGT	2760
TGGAAAACAA AGTCAAAAGT GACACTATAT TCAGCCACAT AGTTTTTATA AGCAGCACTT	2820
CCTTGTCCAA TATCATGATT ATGCTGAATC AAATCGTCTG CCACATAATC ACTCCACTGC	2880
TCTAGCTCCC CATTTTGGAA AATTTCGTGC AAGAAACGGC GAACCAGCTT TTTATTTTCT	2940
GCTTTCTTAT CCAAATCCTT GATTTCAAAA TCTCCAAAAA TTTGATCTAG TTGGTCATTT	3000
TCAGGTGTTT GATAGTAGTC AATGACATCC CAATGCTCAA CAATACAACC ATTCTCATCC	3060
TCACGGAAAG TATCCGTCGT CACCCATTGA GCTTCTCCAC CATTCAGATA TTGATGAACA	3120
TGAACAAAGA CCAGATTGCC ATCCTCAATG GTGCGGACAA TCTTAATCTG ACGCTCTGGA	3180
TGACGCTCAA AGAAATCTGC AAAGAAGGCT GCAAATCCTT CTTTCCCGTC AGGAACACCT	3240
GTGGAATGTT GGATATAGGT ATCCCTTACA GACTGGGCTT GAGCCTCAGC AACTCGTCCG	3300
TCTTGAATGG CATGGATGTA TAGGTTGTGA GCATTTTCA CTTGTTGTGA CATATTCTAA	3360
ACCTCATTTT CCTTCTCTTT CAGATTCGCC AAAATCTTT CTTGAAAACC TTCAAATTGG	3420
TGAATTTCTT CCTCTGAAAA TCCTTTGTAA AAGATAGTAT CCAATTTCTG ACTGACACGA	3480

272

TGCCCCACTT	CTTTCTGGGA	CTTGCCTAAC	TCCGTTAAAA	CTAAATACTT	CTTACGCTTG	3540
TCTTTTCCAC	ACGGACTAAC	AATTACAAGC	TTTTGTTCCT	CTAGCTTTTT	TATCATAGTC	3600
GTCAGCGTAT	TATTGCAAG	TCCAGTCGCA	AGCGCGATAT	CTGTGCGAGT	TGCGCAGCCA	3660
GTTTCACTAT	TCCATAAAAC	CGCTAAAATC	TTGCCCTGTT	CACCCCTATA	AAGAGCCTCA	3720
GGATCTTGAC	TCAGTAACTT	TTGAAAAATC	CGCCCATTC	ACAAACGAAT	ATGATGGGCT	3780
AGCAAATGAC	CATCTTTCAT	AACACCTCCA	ATTTATTTTCG	ATATCGAAAT	GAATAAAACA	3840
ATTGTAACAC	TCATCGTTCT	AACTGTCAAC	TATTTTCGATT	TAGAAATAAT	TTTTGATAAT	3900
TATCCACACC	ACCATACTCC	GGCTCAACTA	ACTTTTAACG	AGAGTTTCTA	AACTCCTTCG	3960
TCCTCCAGTC	TACAAAAGCC	TTCCATTTCGT	ACTATCCTAT	ATTTTATGAG	GGGACACATT	4020
TTTCCTATCA	GACCATTTAT	TTTAAAGATA	GAAGTAAATC	ATAATGCTT	CCATCTGTTC	4080
TTTTATAGTA	TATTGAAGTT	AGACTAGAGC	ACTGTATCTT	CTAAAACATT	GATAGAAAGC	4140
GATTTGAATT	TCCCAATCAA	TTTGTTCGTA	TTTATAGCAT	TTGAAACTG	GAATAGGACA	4200
CCATGACTGC	TAAAAGATTT	CTATAAATTC	ATTTAATTTT	CTCAATCAAT	TTGTTTCATAT	4260
CTTATTTTCAT	TCCGCTATAA	TTTCACCTTA	CCCTATCTTT	TTGCTAGCAC	CCTTCAAACA	4320
GCCTATCCCC	TACCGTTTGA	CGATTCTCTCA	CTTCGCTCCA	CTTCCATTAC	AGAAGTTTCT	4380
TCACTACTAT	GGGCTCGGCT	GACTTCTCAT	GATTCCTTGT	TACTACTATT	TGAACGCTCA	4440
CGAGATAGAT	CTTACAAAAA	ATGCTTTGAT	CCACAATGGA	ATCAAAGCAT	TTTAAAGAGT	4500
TCCTCATACA	TAAGCGCAGA	AGTCGCGAGT	CCTCTGTA	TGGCTTCTTC	TCTTTTGACA	4560
AAGCGAGCCA	AGTTGAGCAA	CTCAGGTGCT	GGATGTTTGG	GATTTAGGAG	CAATTCACGA	4620
TTGACCAGGC	CTGAGAGACG	AACTGCCTGC	AATTGCTCAT	TTGTAGTAGG	CAGTTTTTTA	4680
GTAGTCTCTA	GGAGAGCAGC	AACTAAATCT	TCACTCAAAT	CATGTCGAGC	ATGATTGTAA	4740
AGATCTTTTA	TAAGGCTTTC	TAGGTTTGGT	TCTACCATCC	CTACCACCTC	CCTTATGGTT	4800
TAATAATGTT	TAATCAAATC	AACCGTTGAA	CGATCCAATT	TCTTCACCAA	GGCTTGTAAG	4860
AAAGCTTGCG	CTTCTAGGAA	GTCATCCATT	GCATAGAGGG	TTTGGTGAGA	ATGGATATAA	4920
CGAGCGCAGA	CACCGATAGT	TGTTGATGGG	ACACCACCAT	TTTTCAGATG	AGCTGCACCT	4980
GCATCTGTTT	CGCCTTTACC	ACAGTAGTAT	TGGTACTTGA	TACCAGCTTC	TTCAGCCGTT	5040
GTCAAAAGGA	AATCCTTCAT	CCCTGGGAGA	AGCAAGTGAC	CTGGATCATA	GAAACGAATC	5100
AAGGTTCCAT	CTCCAATCTT	GCCTTGACCA	CCGTAGACAT	CACCTGCTGG	TGAGCAATCA	5160
ACTGCGAGGA	AGACTTCTGG	GTCAAACCTG	GTTGTAGAGG	TATGAGCGCC	ACGCAGACCA	5220
ACTTCTTCTT	GGACGTTAGA	ACCCAGATAG	AGTTCATTGC	CGAGTTTTTG	ACCCGATAAA	5280

273

GCTTCAGCTA GCTCGCTTAC CATGAGGACA CCGTAGCGGT TATCCCAAGC TTTTGAGATG	5340
ATATTTTTTT CATGGCTGT CAAAATTGCA GAACTATCTG GTACAATGGT ATCACCAGGA	5400
CGGATGCCAA AACTTTCTGC CTCAGCCTTG TCCGCAAAAC CACCATCAAA AACGATATCG	5460
GCAATGGCTG GCATGGTTGG TCCCCCTTT CCACGAGTCA AATGCGGAGG AACAGAACCT	5520
GAAATCACAG GAATTTCATG ACCATCACGA GTCAAGAGTT TGAAACGTTG GCTGCTAACC	5580
ACCATGGGGT TCCAGCCACC GATTTCTACG ACACGGAAGG TACCATCTGG CTTGATTTTCG	5640
CTGACCATAA AACCAACTTC GTCCATATGA GAAGCGACCA AGACGCGCGG TGCATCCACA	5700
GCTTCTGAAT GTTTGATACC AAAAATACCA CCAAGCCAT CTGTACCAC TTCATCCACA	5760
TGCGGTGTCA ACTTTTCACG AAGATAAGCA CGACAGGCG CTTCATGACC TGAGACTGCA	5820
GCAAGTTCTG TTACTTCTTT AATTTTGTAA AATAATGTTG TCATTTCACT TCCTTCTTTC	5880
TTTCATCCAT TTTACCACTT TTTATAGGAG AAGGATAGTG GGAAGGTGGA TTTCTAAGTT	5940
AGTATCTTAG TCCTGCTCTA TCTTAGAAAA GGATAGTATT CTCTTGCATG TAGTGCAAAA	6000
TCTAGTAAAC ATTCAAAAT TAACTCGAAT ATTTATTTCC AAACAAAAA ACAATACACC	6060
ATCAAAGTTG TTTGGATTTT TCATGAAATT TACAGAAAAT AGTTGACTTC CCTTTCTTCT	6120
TTCTTTAAAT ATATAGTTGG TTGAGTTTGG AATAGTACGC TGTAGCTGCT AAAACATTTT	6180
TAGAAATTAA TTTGACTTTC CTAATAGAGT TGTTTCATATC TTATTTCAAT TTAATAAGT	6240
ACAAAAC TAG AAAAGGAAAA AATCATGACC AGG	6273

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28171 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

ACAACCTTTT TCAAAAATC ACCTTGTAC GGAGATGTTT TGCTTCTGC TATTATTTTC	60
GGTTATATTC ATATCAATTT TGCTTTAACT CCTCTTGCTT TTTTCATTTA TGCTAGTGGA	120
GGTCTTATTT TAGCTCTATT GTATCGCATG ACTAAAAATC TCTACTATCC AATACTAGTT	180
CATATTCTCA TTAATATCAC TGCCTTCTGG GATGTGTGGT TGCTCCTATT TTCAGGAAGT	240
TAGCTTACTA AAATAATGTC GGAACCTTCC GGCATTTTCT TTTTTCACAA ATAGTCAACG	300
TTTTTCTTTT CGATATTGTA GTGGTGTGTA TCCAGTTATT TTTTGAATT GATTTTGAAA	360

274

ATAAGGTTGA CTTGAGAAAG GCAGATAGTG AAGATAGTTA AGAAGAATAG GATGTTCTTTT	420
TTTCCTTTTT GGAAACTTC TAAAAATATGG TATAATGAAA AGATAAAGAA GTTGGGGTA	480
GAAGATGAAC ATTCAACAAT TACGCTATGT TGTGGCTATT GCCAATAGTG GTACTTTTCG	540
TGAAGCTGCT GAAAAGATGT ATGTTAGTCA GCCGAGTCTG TCTATTTCTG TTCGTGATTT	600
GGAAAAAGAG TTGGGCTTTA AGATTTTCCG TCGGACCAGC TCAGGGACTT TCTTGACCCG	660
TCGTGGGATG GAATTTTATG AAAATCGCA AGAATTGGTT AAAGGATTTG ATATTTTTC	720
AAATCAGTAT GCCAATCCTG AAGAAGAAA AGATGAATTT TCTGTTGCTA GCCAGCACTA	780
TGACTTCTTG CCACCAACTA TTACGGCCTT TTCAGAGCGC TATCCTGACT ATAAGAACTT	840
CCGTATTTTT GAATCAACTA CTGTTCAAAT ATTAGATGAA GTGGCGCAAG GGCATAGTGA	900
GATTGGGATT ATCTACCTCA ACAATCAAAA TAAAAAGGGG ATTATGCAAC GGGTTGAAAA	960
ATTAGGTCTG GAGGTCATCG AATTGATTCC TTTCCATACC CATATTTATC TCCGTGAGGG	1020
TCATCCTTTA GCCCAGAAAG AGGAATTAGT CATGGAGGAT TTAGCGGATT TACCAACGGT	1080
TCGTTTCACT CAAGAGAAAG ACGAGTACCT TTATTATTCA GAGAACTTTG TCGATACCAG	1140
CGCTAGCTCA CAGATGTTTA ATGTGACAGA CCGTGCCACC TTGAATGGTA TTTTGGAGCG	1200
GACGGACGCC TATGCGACAG GTTCTGGATT TTTAGATAGT GACAGTGTTA ATGGCATTAC	1260
AGTTATTCGT CTCAAGGATA ACCTAGATAA CCGCATGGTC TATGTTAAAC GTGAAGAAGT	1320
GGAGCTTAGT CAAGCTGGGA CTCTCTTCGT AGAAGTCATG CAAGAATATT TTGATCAAAA	1380
GAGGAAATCA TGAAAAAAG AGCAATAGTG GCAGTCATTG TACTGCTTTT GATTGGGCTG	1440
GATCAGTTGG TCAAAATCCTA TATCGTCCAG CAGATTCCAC TGGGTGAAGT GCGCTCCTGG	1500
ATCCCCAATT TCGTTAGCTT GACCTACCTG CAAAATCGAG GTGCAGCCTT TTCTATCTTA	1560
CAAGATCAGC AGCTGTTATT CGCTGTCATT ACTCTGGTTG TCGTGATAGG TGCCATTTGG	1620
TATTTACATA AACACATGGA GGAATCATTC TGGATGGTCT TGGGTTTGAC TCTAATAATC	1680
GCGGGTGGTC TTGGAACTT TATTGACAGG GTCAGTCAGG GCTTTGTTGT GGATATGTTT	1740
CACCTTGACT TTATCAACTT TGCAATTTTC AATGTGGCAG ATAGCTATCT GACGGTTGGA	1800
GTGATTATTT TATTGATTGC AATGCTAAAA GAGGAAATAA ATGGAAATTA AAATTGAAAC	1860
TGGTGGTCTG CGTTTGATA AGGCTTTGTC AGATTTGTCA GAATTATCAC GTAGTCTCGC	1920
GAATGAACAA ATTAAATCAG GCCAGGTCTT GGTCAATGGT CAAGTCAAGA AAGCTAAATA	1980
CACAGTCCAA GAGGGTGATG TCGTCACTTA CCATGTGCCA GAACCAGAGG TATTAGAGTA	2040
TGTGGCTGAG GATCTTCCGC TAGAAATAGT CTACCAAGAT GAGGATGTGG CTGTCGTTAA	2100
CAACCTCAG GGAATGGTTG TGCACCCGAG TGCTGGTCAT ACCAGTGGA CCCTAGTAAA	2160

275

TGCCCTCATG TATCATATTA AGGACTTGTC GGGTATCAAT GGGGTTCGTC GTCCAGGGAT	2220
TGTTACACCGT ATTGATAAGG ATACGTCAGG TCTTCTCATG ATTGCTAAAA ACGATGATGC	2280
GCATCTAGCA CTGCCCCAAG AACTCAAGGA TAAAAAGTCT CTCCGCAAAT ATTGGGCGAT	2340
TGTTTCATGGA AATCTACCTA ATGATCGTGG TGTAATTGAA GCGCCGATTG GCCGGAGTGA	2400
AAAAGACCGT AAGAAACAGG CTGTAAGTGC TAAAGGGAAG CCTGCAGTGA CGCGTTTTC	2460
CGTCTTGGA CGCTTTGGCG ATTATAGCTT AGTAGAGTTG CAACTGGAGA CAGGGCGCAC	2520
TCATCAAATC CGTGTCCACA TGGCTTATAT CGGCCATCCA GTCGCTGGTG ATGAGGTCTA	2580
TGGTCCTCGC AAGACTTTGA AAGGACATGG ACAATTTCTT CATGCCAAGA CTTTAGGTTT	2640
TACTCATCCG AGAACAGGTA AGACCTTGGA ATTTAAAGCA GATATCCCAG AGATTTTAA	2700
GGAAACCTTG GAGAGATTGA GAAAGTAAGA ATGAAAAAGA AATTAAGTAG TTAGCACTT	2760
GTAGGCGCTT TTTTAGGTTT GTCATGGTAT GGGAAATGTC AGGCTCAAGA AAGTTCAGGA	2820
AATAAAATCC ACTTTATCAA TGTTCAGAA GGTGGCAGTG ATGCGATTAT TCTTGAAAGC	2880
AATGGACATT TTGCCATGGT GGATACAGGA GAAGATTATG ATTTCCAGA TGGAAGTGAT	2940
TCTCGCTATC CATGGAGAGA AGGAATTGAA ACGTCTTATA AGCATGTTCT AACAGACCGT	3000
GTCTTTCGTC GTTTGAAGGA ATTGGGTGTC CAAAACTTG ATTTTATTTT GGTGACCCAT	3060
ACCCACAGTG ATCATATTGG AAATGTTGAT GAATTACTGT CTACCTATCC AGTTGACCGA	3120
GTCTATCTTA AGAAATATAG TGATAGTCGT ATTACTAATT CTGAACGTCT ATGGGATAAT	3180
CTGTATGGCT ATGATAAGGT TTTACAGACT GCTGCAGAAA AAGGTGTTTC AGTTATTCAA	3240
AATATCACAC AAGGGGATGC TCATTTTCAG TTGTTGGGACA TGGATATTCA GCTCTATAAT	3300
TATGAAAAATG AAAGTGATTC ATCGGGTGAA TTAAAGAAAA TTTGGGATGA CAATTCCAAT	3360
TCCTTGATTA GCGTGGTGAA AGTCAATGGC AAGAAAAATTT ACCTTGGGGG CGATTTAGAT	3420
AATGTTTCATG GAGCAGAAGA CAAGTATGGT CCTCTCATTG GAAAAGTTGA TTTGATGAAG	3480
TTTAATCATC ACCATGATAC CAACAAATCA AATACCAAGG ATTTTCATTAA AAATTTGAGT	3540
CCGAGTTTGA TTGTTCAAAC TTCGGATAGT CTACCTTGA AAAATGGTGT TGATAGTGAG	3600
TATGTTAATT GGCTCAAAGA ACGAGGAATT GAGAGAATCA ACGCAGCCAG CAAAGACTAT	3660
GATGCAACAG TTTTGTATAT TCGAAAAGAC GGTTTTGTCA ATATTTCAAC ATCCTACAAG	3720
CCGATTCCAA GTTTTCAAGC TGGTTGGCAT AAGAGTGCAT ATGGGAACTG GTGGTATCAA	3780
GCGCCTGATT CTACAGGAGA GTATGCTGTC GGTGGAATG AAATCGAAGG TGAATGGTAT	3840
TACTTTAACC AAACGGGTAT CTTGTTACAG AATCAATGGA AAAAATGGAA CAATCATGG	3900

276

TTCTATTTGA CAGACTCTGG TGCTTCTGCT AAAAAATTGGA AGAAAATCGC TGGAAATCTGG	3960
TATTATTTTA ACAAAGAAAA CCAGATGGAA ATTGGTTGGA TTCAAGATAA AGAGCAGTGG	4020
TATTATTTGG ATGTTGATGG TTCTATGAAG ACAGGATGGC TTCAATATAT GGGGCAATGG	4080
TATTACTTTG CTCCATCAGG GGAAATGAAA ATGGGCTGGG TAAAAGATAA AGAAACCTGG	4140
TACTATATGG ATTCTACTGG TGTCATGAAG ACAGGTGAGA TAGAAGTTGC TGGTCAACAT	4200
TATTATCTGG AAGATTCAGG AGCTATGAAG CAAGGCTGGC ATAAAAAGGC AAATGATTGG	4260
TATTTCTACA AGACAGACGG TTCACGAGCT GTGGGTTGGA TCAAGGACAA GGATAAATGG	4320
TACTTCTTGA AAGAAAATGG TCAATTACTT GTGAACGGTA AGACACCAGA AGGTTTACT	4380
GTGGATTCAA GTGGTGCCTG GTTAGTGGAT GTTTCGATCG AGAAATCTGC TACAATTAAA	4440
ACTACAAGTC ATTCAGAAAT AAAAGAATCC AAAGAAGTAG TGAAAAAGGA TCTTGAAAAT	4500
AAAGAAACGA GTCAACATGA AAGTGTACA AATTTTCAA CTAGTCAAGA TTTGACATCC	4560
TCAACTTCAC AAAGCTCTGA AACGAGTGTA AACAAATCGG AATCAGAACA GTAGTAGAAA	4620
AGAAGGTTT AGGGCCTTCT TTTTCCTATC AACTCTTTTC TATTTCTGT TATTCAATGT	4680
ATAATGGATA AATATGAATA ATCGGAGTGA GACTATGAAA TACAAACGGA TTGTCTTTAA	4740
GGTGGGTACT TCITCTCTGA CAAATGAGGA TGGAAGTTTA TCACGTAGTA AGGTAAAGGA	4800
TATTACCCAG CAGTTGGCTA TGCTGCACGA GGCTGGTCAT GAGTTGATTT TGGTGTCTTC	4860
AGGTGCCATT GCGGCTGGTT TTGGAGCCTT AGGATTTAAA AAGCGTCCGA CTAAGATTGC	4920
TGATAAACAG GCTTCAGCAG CGGTAGGGCA GGGGCTTTTG TTGGAAGAAT ATACAACCAA	4980
TCTTCTCTTG CGTCAAATCG TTTCTGCACA AATCTTGCTG ACCCAAGATG ACTTTGTGGA	5040
TAAGCGTCGT TATAAAAATG CCCATCAGGC TTTGTGCGTT TTGCTCAACC GTGGGCAAT	5100
TCCTATCATC AATGAGAATG ATAGTGTCTG TATTGATGAG CTCAAGGTTG GGGACAATGA	5160
CACTCTAAGT GCTCAAGTAG CGGCGATGGT CCAAGCAGAC CTTTGTAGTT TCTTGACAGA	5220
TGTGGACGGT CTCTATACTG GAAATCCTAA TTCAGATCCA AGAGCCAAAC GCTTGGAGAG	5280
AATCGAGACC ATCAATCGTG AGATTATTGA TATGGCTGGT GGAGCTGGTT CGTCAAACGG	5340
AACTGGGGT ATGTTAACC AATCAAGGC TGCAACTATC GCGACGGAAT CAGGAGTTCC	5400
TGTTTATATC TGCTCATCCT TGAAATCAGA TTCCATGATT GAGGCGGCAG AGGAGACCGA	5460
GGATGGTTCT TACTTTGTTG CTCAAGAGAA GGGGCTTCGT ACCCAGAAAC AATGGCTTGC	5520
CTTCTATGCT CAGAGTCAAG GTTCTATTG GGTGATAAA GGGGCTGCGG AAGCTCTCTC	5580
TCAATATGGA AAGAGCTTC TCTTATCTGG TATCGTTGAA GCAGAAGGAG TCTTTTCTTA	5640
CGGTGATATC GTGACAGTAT TTGACAAGGA AAGTGAAAA TCACTTGGA AAGGACCGCT	5700

277

GCAATTTGGA GCATCTGCTT TGGAGGATAT GTFGCGTTCT CAAAAAGCCA AGGGTGTCTT	5760
GATTTACCGT GACGACTGGA TTTCCATTAC TCCTGAAATC CAACTACTTT TTACAGAATT	5820
TTAGAGGTAA ACTATGGTGA GTAGACAAGA ACAATTTGAA CAGGTACAGG CTGTTAAAAA	5880
ATCGATTAAC ACAGCTAGTG AAGAAGTGAA AAACCAAGCC TTGCTAGCCA TGGCTGATCA	5940
CTTAGTGGCT GCTACTGAGG AAATTTTAGC GGCTAATGCC CTCGATATGG CAGCGGCTAA	6000
GGGGAAAAATC TCAGATGTGA TGTGGATCG TCTTTATTTG GATGCAGATC GTATAGAAGC	6060
GATGGCAAGA GGAATTCGTG AAGTGGTTGC CTTACCAGAT CCAATCGGTG AAGTTTTAGA	6120
AACAAGTCAG CTTGAAAATG GTTTGGTTAT CACAAAAAA CGTGTAGCTA TGGGTGTCAT	6180
CGGTATTATC TATGAAAGCC GTCCAAATGT GACGTCTGAT GCGGCTGCTT TGACTCTTAA	6240
GAGTGGAAAT GCGGTTGTTT TTCGTAGTGG TAAGGATGCC TATCAAACAA CCCATGCCAT	6300
TGTCACAGCC TTGAAGAAGG GCTTGGAGAC GACTACTATT CATCCAAATG TGATTCAACT	6360
GGTGGAGGAT ACTAGCCGTG AAAGTAGTTA TGCTATGATG AAGGCCAAGG GCTATCTAGA	6420
CCTTCTCATT CCTCGTGGAG GAGCTGGCTT GATCAATGCA GTGGTTGAGA ATGCGATTGT	6480
ACCTGTTATC GAGACAGGGA CTGGGATTGT CCATGTCTAT GTGGATAAGG ATGCAGACGA	6540
AGACAAGGCG CTGTCTATCA TCAACAATGC TAAAACCACT CGTCCTTCTG TTTGTAATGC	6600
CATGGAGGTT CTGCTGGTTC ATGAAAACAA GGCAGCAAGC TTCCTTCCTC GCTTGGAGCA	6660
AGTGTGGTTC GCAGAGCGTA AGGAAGCTGG ACTGGAACCA ATTCAATTCC GCCTAGATAG	6720
CAAAGCAAGC CAGTTTGTTC CAGGTCAAGC AGCTGAGACC CAAGACTTTG ACACCGAGTT	6780
TTTAGACTAT GTCCTTGCTG TTAAGGTTGT GAGCAGTTTA GAAGAAGCGG TTGCGCACAT	6840
TGAATCCAC AGCACCCATC ATTCGGATGC TATTGTGACG GAAAATGCTG AAGCTGCAGC	6900
ATACTTTACA GATCAAGTGG ACTCTGCAGC GGTGTATGTT AATGCCTCAA CTCGTTTCAC	6960
AGATGGAGGA CAATTTGGTC TTGGTTGTGA AATGGGATT TCTACTCAGA AATTGCACCC	7020
GCGTGGTCCC ATGGGCTTGA AAGAGTTGAC CAGCTACAAG TATGTGGTTG CCGGTGATGG	7080
GCAGATAAGG GAGTAAGAGA TGAAGATTGG ATTTATCGGT TTGGGGAATA TGGGTGCTAG	7140
CTTGGCAAAA TCTGTCTTGC AGACTAGGAC GTCAGATGAG ATTCTCCTTG CCAATCGTAG	7200
TCAAGCTAAG GTAGATGCTT TCATTGCAGA CTTTGGTGGT CAGGCTTCCA GCAATGAAGA	7260
AATGTTTGCA GAAGCAGATG TGATTTTCT AGGAGTTAAG CCTGCTCAGT TTTCTGAACT	7320
GCTTCTCAA TACCAGACCA TCCTTGAAAA AAGAGAAAGT CTCTCTTTGA TTTGATGGC	7380
AGCTGGATTG ACCTTAGAAA AACTAGCAAG TCTTATCCCA AGTCAACACC GAATTATTCG	7440

278					
TATGATGCCT	AATACCCCTG	CTTCTATCGG	GCAAGGAGTG	ATTAGTTATG	CCTTGTCTCC 7500
TAATTGCAGG	GCTGAGGACA	GTGAGCTCTT	TTATCAGCTT	TTAGCCAAGG	CTGGTCTCTT 7560
GGTTGAACTA	GGAGAAAGTT	TAATCGATGC	AGCGACAGGT	CTTGCAGGTT	GTGGACCAGC 7620
CTTTGTCTAT	CTTTTATCG	AGGCCTTGGC	AGATGCAGGT	GTTTCAGACAG	GATTACCACG 7680
AGAAATAGCA	TTGAAAATGG	CAGCACAAAC	TGTGGTAGGA	GCTGGGCAAT	TGCTCCTTGA 7740
AAGTCAGCAA	CATCCTGGAG	TATTGAAAGA	CCAAGTCTGT	AGCCCAGGCG	GTTGCGACTAT 7800
CGCTGGTGTA	GCAAGCCTAG	AAGCGCATGC	TTTCCGAGGA	ACAGTCATGG	ATGCAGTTCA 7860
TCAAGCCTAC	AAACGAACAC	AAGAAGTAGG	TAAATAAGAG	GTAGTTTGA	CTGCCTCTTT 7920
TATGGTGGCT	GAAATGAGAA	GACACAAAAA	GATTGTCACA	AACCCCTATT	TTTTTGATAG 7980
AATAGAAGTA	GTAAAAAAGA	AATGAGTTAG	ACATGTCAAA	AGGATTTTGA	GTCTCTCTTG 8040
AGGGACCAGA	GGGAGCAGGC	AAGACCAAGT	TTTTAGAGGC	TCTGCTACCA	ATTTTAGAGG 8100
AAAAAGGAGT	AGAGGTGTG	ACGACCCGTG	AACCTGGCGG	AGTCTTGATT	GGGGAGAAGA 8160
TTGGGGAAGT	GATTTTGAT	CCAAGTCATA	CTCAGATGGA	TGCTAAAACA	GAGCTACTTC 8220
TCTATATTGC	CAGTCGCAGA	CAGCATTTGG	TGGAAAAAGT	TCTTCCAGCC	CTTGAAGCTG 8280
GCAAGTTGGT	CATCATGGAT	CGTTTATCG	ATAGTCTGT	TGCCTATCAG	GGATTTGGTC 8340
GTGGCTTAGA	TATTGAAGCC	ATTGACTGGC	TCAATCAGTT	TGCGACAGAT	GGCCTCAAAC 8400
CCGATTTGAC	ACTCTATTTT	GACATCGAGG	TGGAAGAAGG	GCTGGCTCGT	ATTGCTGCTA 8460
ATAGTGACCG	CGAGGTTAAT	CGTTTGGATT	TGGAAGGGTT	GGACTTGCA	AAAAAAGTTC 8520
GTCAAGGCTA	CCTTCTCTTT	CTGGATAAAG	AGGGAAATCG	CATTGTCAAG	ATTGATGCTA 8580
GTCTCCCTTT	GGAGCAAGTT	GTGGAACATA	CCAAGGCTGT	CTTGTTTGAC	GGAATGGGCT 8640
TGGCCAAATG	AAACAAGATC	AACTAAAGGC	TTGGCAACCA	GCTCAGTTTG	ACCGTTTGT 8700
CCGTATCTTA	GAACAAGACC	AGCTCAATCA	CGCCTATCTC	TTTTCAGGTT	TCTTTGAAAG 8760
CTTGGAATG	GCGCAATTTT	TAGCTAAGAG	CCTCTTTTGT	ACGGATAAAG	TGGCGTCTT 8820
ACCATGTGAG	AAATGCCGAA	GTTGCAAGCT	GATTGAACAG	GGAGAATTTT	CCGATGTCAC 8880
CTTGATTAAA	CCAGTTAATC	AGGTCATTAA	GACGGAACGC	ATTGAGAAAT	TGGTGGGTCA 8940
GTTTCTCAA	GCAGGGATTG	AAAGCCAGCA	ACAGGCTTTT	ATCATCGAGC	AAGCGGATAA 9000
AATGCATCCC	AACGCAGCCA	ATTCTCTGCT	CAAGGTCATC	GAAGAACCCC	AGAGTGAAGT 9060
TTATATTTTC	TTCTTGACTA	GCGATGAGGA	AAAGATGTTA	CCGACAATCC	GAAGTCGGAC 9120
TCAGATCTTC	CACTTTAAAA	AGCAAGAAGA	AAAACCTATC	TTACTCTTAG	AACAAATGGG 9180
ACTTGTTAAG	AAAAAAGCGA	CTCTTTTAGC	TAAGTTTAGT	CAATCGCGAG	CTGAAGCAGA 9240

AAAGTTGGCT AATCAGGCAA GTTTTGGAC CTGGTCGAT GAAAGTGAAC GCCTGCTGAC	9300
TTGCTTAGTA GCTAAGAAAA AAGAAAGTTA TCTACAGGTT GCCAAATTAG CCAACTTGGC	9360
AGATGATAAG GAAAAACAGG ATCAGGTTTT ACGGATTCTT GAAGTTCTCT GTGGGCAGGA	9420
CCTCTTGACG GTAAGAGTAA GAGTGATTCT ACAAGATTTA CTAGAAGCTA GAAAAATGTG	9480
GCAAGCTAAT GTCAGCTTTC AAAATGCCAT GGAATATCTG GTCTTGAAAG AAATATAAAC	9540
TCAAAAATGA ATGATAAAGA AAGGAAAGGG CTGTTTTATG GACAAAAAAG AATTATTTGA	9600
CGCGCTGGAT GATTTTCC CACAATTATT GGTAACCTTA GCCGATGTGG AAGCCATCAA	9660
GAAAAATCTC AAGAGCCTGG TAGAGGAAAA TACAGCTCTT CGCTTGAAA ATAGTAAGTT	9720
GCGAGAACGC TTGGGTGAGG TGGAAGCAGA TGCTCCTGTC AAGGCCAAGC ATGTTCTGTA	9780
AAGTGTCGGT CGCATTTACC GTGATGGATT TCACGTATGT AATGATTTTT ATGGACAACG	9840
TCGAGAGCAG GACGAGGAAT GTATGTTTTG TGACGAGTTG CTATACAGGG AGTAGGCATG	9900
CAGATTCAAA AAAGTTTAA GGGCAGTCT CCCTATGGCA AGCTGTATCT AGTGGCAACG	9960
CCGATTGGCA ATCTAGATGA TATGACTTTT CGTGCTATCC AGACCTTGAA AGAAGTGGAC	10020
TGGATTGCTG CTGAGGATAC GCGCAATACA GGGCTTTTGC TCAAGCATT TGACATTTCC	10080
ACCAAGCAGA TCAGTTTTC TGAGCACAAT GCCAAGGAAA AAATTCCTGA TTTGATTGGT	10140
TTCTTGAAAG CAGGGCAAAG TATTGCTCAG GTCTCTGATG CCGGTTTGCC TAGCATTTCA	10200
GACCTGGTC ATGATTTAGT TAAGGCAGCT ATTGAGGAAG AAATTGCAGT TGTGACAGTT	10260
CCAGGTGCCT CTGCAGGAAT TTCTGCCTTG ATTGCCAGTG GTTTAGCGCC ACAGCCACAT	10320
ATCTTTTACG GTTTTACC GAGAAAATCA GGTGAGCAGA AGCAATTTTT TGGCTTGAAA	10380
AAAGATTATC CTGAAACACA GATTTTTTAT GAATCACCTC ATCGTGTAGC AGACACGTTG	10440
GAAAATATGT TAGAAGTCTA CGGTGACCGC TCCGTTGTCT TGGTCAGGGA ATTGACCAAA	10500
ATCTATGAAG AATACCAACG AGGTACTATC TCTGAGTTAT TAGAAAGCAT TGCTGAAACG	10560
CCACTCAAGG GCGAATGTCT TCTCATGTT GAGGGTGCCA GTCAGGGTGT GGAGGAAAAG	10620
GACGAGGAAG ACTTGTTCTG AGAAATTC AAACCGCATCC AGCAAGGTGT GAAGAAAAAC	10680
CAAGCTATCA AGGAAGTCGC TAAGATTTAC CAGTGGAAATA AAAGTCAGCT CTACGCTGCC	10740
TACCACGACT GGAAGAAAA ACAATAAAGG GAGACAGGAT GTAATAATTC TGTCTGTTTC	10800
TGTTTAACTT AATTAGTGAT GATAATATAA AGATGTATCA CTTGGTATAG AAGCTTTGGT	10860
ATTAAGTTTT TTATTAAGCC CATACGGAAT ACCGATGGTT GGAGCAGCAG TTATAGCGTT	10920
CTTAGAAGGT ATAAATAGAA AAATAAGGTC ATTTTAAATC AAAGGATTGA TAAATCAGAA	10980

280

AGAAGGTGAT TTTTTCGCAA CATACGAAAA TAAAGAAGAA CTAAAAGCTG AGATAGAGAA	11040
AACATTTGAG AAATATATTT TAGAATTTGA TAATATTCCA GAAAATTTAA AAGATAAGAG	11100
AGCTGATGAA GTTGACAGAA CTCCAGCAGA AAACCTTTCCT TATCAGGTTG GTTGGACCAA	11160
CTTGCTTCTT AAATGGGAAG AAGATGAAAG AAAGGGGCTT CAAGTAAAAA CACCATCGGA	11220
TAAATTTTAA TGGAAATCAAC TTGGTGAATT ATATCAGTGG TTCACAGATA CCTACGCTCA	11280
TTTATCTCTG CAAGAGTTGA AAGCAAAATT AAATGAAAAT ATTAATTCTA TCTCTGCAAT	11340
GATTGATTCTG TTGAGTGAGG AAGAATTATT TGAACCGCAT ATGAGAAAGT GGGCTGATGA	11400
AGCGACTAAA ACAGCGACTT GGGAAAGTGA TAAGTTTATT CATGTAAATA CGGTTGCACC	11460
TTTTGGAAGT TTCAGAACTA AAATCAGAAA ATGGAAGAAG ATAGTATTAT AAATTATATT	11520
TTTAACCTTA AAAAAATTTCA TAAAAATGGT TACCAAAGGC GATAGAAGAA AACTATCGT	11580
CTTTTCTTTT GCAAATTTTT AAGAAGGGAG GTGATCTTGC ATGGACTTTG AATATTTTAA	11640
TAACAGAGAA GCGGAAAGAT TTAACCTCTT AAAAGTACCG GAGATATTAG TTGATAGAGA	11700
AGAATTTCTG GGCTTATCAG CAGAAGCAAT TATCCTTTAT TCCATACTTC TTAACAGAC	11760
AGGAATGTCA TTTAAGAATA ACTGGATAGA CAAGGAAGGC AGAGTATTTA TCTATTTTAC	11820
TGTCGAAGAA ATTATGAAAA GAAGAAATAT CTCAAAGCCA ACTGCCATAA AAACATTAGA	11880
TGAGCTTGAT GTAAAAAAGG AATAGGACTG ATCGAAAGAG TAAGGCTTGG ACTTGGTAAG	11940
CCGAACATCA TTTATGTTAA AGACTTTATG AGTATATTTT AGGTAAAAGA AAATGACTTA	12000
CAGAAGTCAA AAAACTTAAC TTCAGAAGTA AAAGATTTTA ACCTCAGAAG TAAAGAAAAT	12060
GAAGTTCAAG AGGTAAAGAA CCTTGACTCT AACTATATAG AGAATAATAA GAGTAAGTAT	12120
AGTAAGAGAG AATATAGTTT TGGTGAAAAC GGACTTGGA CATTTCAAAA TGTGTTTTTA	12180
GCTGCTGAAG ATATATCGGA TTTACAAATC ATAATGAAGT CACAGCTTGA GAATTACATT	12240
AGACTTCCTG CAAAACCTAG ATCCTAGTTC ATGATTGATA ATGCCAGCAA TCAAATTCAT	12300
TCGTAATCCG AAGCGTTTAC GATGATTTCT ATAGATTGTT GAAAACATTT TAAACGTTTT	12360
TACTTTGGCA AAGATGTTCT CAATCTTGCT TCTCTCCTTG GATAGCGCAT GGTACAGGC	12420
TTTATCTTCA GCTGTTAGCG GCTTGAGTTT GCTGGATTTA CGTGGAGTTT GTACTTGAGG	12480
ATATATCTTC ATGAGCCCTT GATAACCACT GTCAGACAAG ATTTTACCAG CTTGTCCGAT	12540
ATTTCTGCGA CTCATTTTGA ACAACTTCAT ATCAGACAAA TAGTTCACAG CGATATCCAA	12600
AGAAACAATT CTCCCTTGAC TTGTGACAAT CGCTTGAGCC TTCATAGCGT GAAATTTCTT	12660
TTTACCAGAA TGATTCGCTA ATTCTTTTTT TAGGGCGATT GATTTTACT TCCGTCGCAT	12720
CAATCATTAC CGTGTCTTCA GAACTGAGAG GAGTTCTTGA AATCGTAACA CCACTTTGAA	12780

CAAGAGTTAC TTCAACCCAT TGGCTCCGAC GGATTAAGTT GCTTTCGTGA ATACCAAAAT	12840
CAGCCGCAAT TTGTTCATAA GTTCGATATT CTCGCACATA TTGAAGAGTG GCCATAAGAA	12900
GGTCTTCTAG GCTTAATTTA GGTTCGTC CACCTTTTGC GTGTTTAAGT TGATAAGCTG	12960
TTTTTAATAC AGCTAATATC TCTTCAAAAG TCGTGCCTG AACACCAACA AGACGCTTAA	13020
ATCGTGCATC AGTTAGTTGT TTACTTGCTT CATCATTCAT AGAACTACTA TACCATATTT	13080
TGTTTCGCAG GAAGTCTATT GGAAAGTAAG AAATATTGAA GCTGAGGCTA TTAGAAGAAA	13140
TTGTGAGCGT GGTGCTATTT TTTCAGGTAA AATAAATAT CACGAAGATT CACAGTTTAA	13200
AGGAGATCAC TATGTTGAAT GTTATGCTGT TTTAGATAAT ACGGTTATAG CAAGAGATAG	13260
AATAACAGTC CCTATCGATC CGTTATGTGG AAAAGATTTT ATAGAGTAGC ATATAATTGA	13320
TTCTTAACTG GAATACTCAC TATCTCTTTA CATCAAGAAA ATGACTAAAC AGGGAAGTTT	13380
GCCTTCTTCC CTTTTTTTGT TATACTAGTA GAAGAAAAA TTAGAAAGAT TTGTGGGTGT	13440
CAAACAGCCC AGTGGGTGT TTTAATATGG ACTTAGGTCC CACCCAAAGA GGTATTAGTG	13500
TCGTGTCTCA ATCTTATATC AATGTTATCG GTGCTGTTT GGCAGGTTCT GAAGCAGCTT	13560
ACCAAATCGC AGAGCGTGGT ATTCCAGTTA AACTATATGA AATGCGTGGT GTCAAGTCTA	13620
CACCCAGCA TAAACAGAC AATTTTGCTG AGTTGGTTTG TTCCAATTCT TTGCGTGGG	13680
ATGCTTTGAC AAATGCAGTT GGTCTTCTCA AGGAAGAAAT GCGTCGCTTG GGTCTGTTA	13740
TCTTGAATC TGCTGAGGCT ACACGTGTTT CTGCAGGTGG TGCCCTTGCA GTGGACCGTG	13800
ATGGTTTCTC TCAAATGGTG ACCGAAAAG TTGCCAACCA CCCCTTGATT GAAGTGGTTC	13860
GTGATGAAAT TACAGAAATG CCGACAGATG TTATTACGGT TATCGCTACT GGTCCTTTGA	13920
CAAGTGATGC CTTGGCTGAA AAGATTCATG CTCTTAATGA CCGTGCTGGT TTTTATTTCT	13980
ACGATGCGGC AGCGCCTATT ATCGATGTCA AACTATCGA TATGAGCAAG GTCTACCTCA	14040
AATCACGTTA TGATAAGGGA GAAGCGCCT ACCTCAATGC CCCTATGACC AAGCAAGAAT	14100
TTATGGATTT CCATGAAGCT TTGGTCAATG CAGAAGAAGC ACCGCTTAGT TCTTTTGA	14160
AAGAAAAGTA CTTTGAAGGA TGTATGCCTA TCGAAGTCAT GGCCAAACGT GGCATTAAAA	14220
CTATGCTTTA TGGCCCTATG AAGCCAGTCG GTCTTGAGTA CCCAGACGAC TATACAGGAC	14280
CTCGTGATGG AGAATTTAAA ACACCTTATG CGGTTGTGCA ACTTCGTCAG GATAATGCAG	14340
CTGGTAGCCT CTACAATATT GTTGGTTTCC AGACCCACCT CAAATGGGGA GAACAAAAGC	14400
GTGCTTCCA AATGATCCG GGTCTTGAAT ATGCGGAGTT TGTCCGTTAT GGTGTGATGC	14460
ATCGCAATTC TTACATGGAT TCACCAAATC TTCTTGAGCA GACTTACCGT TCTAAGAAAC	14520

282
AACCAAATCT CTTCTTTGCT GGTCAAATGA CCGGTGTGGA AGCCTATGTT GAGTCGGCGG 14580
CTTCAGGCTT AGTGCGGGA ATTAACGCAG CTCGTCTCTT CAAGGAAGAA AGCGAGGCTA 14640
TTTTCCCCGA GACGACAGCG ATTGGAAGCT TAGCTCATTA CATTACCCAT GCCGACAGCA 14700
AACATTTCCA ACCAATGAAT GTCAATTTG GGATCATCAA GGAGTTGGAA GGCGAGCGTA 14760
TCCGTGATAA GAAGGCTCGT TATGAAAAA TTGCAGAGCG TGCCCTTGCC GACTTAGAGG 14820
AATTTTGTGAC TGTCTAATTT TTTTGAAAGA ATTGCTCATG ATACTATAAA AATCTTAGAA 14880
ATTGTGATAA AATAGGTAGG ATGAAAGAAG GAGAGTGAAA ATGGCGAATC CCAAGTATAA 14940
ACGTATTTTA ATCAAGTTAT CAGGTGAAGC CCTTGCCGGT GAACGTGGCG TAGGGATTGA 15000
TATCCAAACA GTTCAAACAA TCGCAAAAGA GATTCAAGAA GTTCATAGCT TAGGTATCGA 15060
AATTGCCCTT GTTATCGGTG GAGGAAATCT CTGGCGTGA GAACCTGCAG CAGAAGCAGG 15120
TATGGACCGT GTTCAGGCAG ATTACACAGG AATGCTTGGG ACTGTTATGA ATGCTCTTGT 15180
GATGGCAGAT TCATTGCAAC AAGTTGGGGT TGATACGCGT GTACAAACAG CTATTGCCAT 15240
GCAACAAGTG GCAGAGCCTT ATGTCCGTGG ACGTGCCCTT CGTCACCTTG AAAAAGGCCG 15300
TATCGTTATC TTTGGTGCTG GAATTGGTTC ACCTTACTTC TCGACAGATA CAACAGCGGC 15360
CCTTCGTGCA GCTGAAATCG AAGCAGATGC CATCCTCATG GCTAAAAATG GTGTCGATGG 15420
TGTTTACAAT GCCGATCCTA AGAAAGATAA GACAGCTGTT AAGTTTGAAG AATTGACCCA 15480
CCGTGACGTT ATCAATAAAG GTCTTCGTAT CATGGACTCA ACAGCTTCAA CCCTCTCAAT 15540
GGACAACGAC ATTGACTTGG TTGTATTCAA CATGAACCAA CCAGGCAACA TCAAACGTGT 15600
CGTATTTGGT GAAAATATCG GAACACAGT TTCAAATAAT ATCGAAGAAA AGGAATAAGA 15660
AAGAATATGG CTAACGCAAT TATTGAAAAA GCTAAAGAGA GAATGACCCA GTCTCACCAA 15720
TCACTTGCTC GTGAATTTGG TGGTATCCGT GCTGGTCGTG CCAATGCAAG CTTGCTTGAC 15780
CGTGACATG TAGAATACTA TGGAGTCGAA ACTCCTCTTA ACCAAATCGC TTCAATTACG 15840
ATTCCAGAAG CGCGTGTTTT GTTGGTAACA CCATTGACA AGTCTTCATT GAAAGACATC 15900
GAACGTGCCT TGAACGCTTC TGATATTGGT ATCACACCGG CTAATGACGG TTCTGTGATT 15960
CGCTTGGTTA TCCCAGCTCT TACAGAAGAA ACTCGTCGTG ACCTTGCTAA AGAAGTGAAG 16020
AAGGTCGGCG AAAATGCTAA AGTGGCTGTC CGCAATATCC GTCGCGATGC TATGGACGAA 16080
GCTAAGAAAC GAGAAAAAGC AAAAGAAATC ACTGAAGACG AATTGAAGAC TCTTGAAAAA 16140
GACATTCAAA AAGTAACAGA CGATGCTGTT AAACACATCG ACGACATGAC TGCTAACAAA 16200
GAGAAAGAAC TTTTGGAAGT CTAATAATAA ACAGAAAAAC TCAGTTGGCA TTGCTGGCTG 16260
AGTTTATTC GAAAGAAGGA AATATGAATA CAAATCTTGC AAGTTTATC GTTGGACTGA 16320

TCATCGATGA AAACGACCGT TTTTACTTTG TGCAAAAGGA TGGTCAAACC TATGCTCTTG	16380
CTAAGGAAGA AGGCCAACAT ACAGTAGGGG ATACGGTCAA AGGTTTTGCA TACACGGATA	16440
TGAAGCAAAA ACTCCGCCTG ACAACCTTAG AAGTGACTGC CACTCAGGAC CAATTTGGTT	16500
GGGGACGTGT CACAGAGGTT CGTAAGGACT TGGGTGTCTT TGTGGATACA GGCCTTCCTG	16560
ACAAGGAAAT CGTTGTGTCA CTCGATATTC TCCCTGAGCT CAAGGAACTC TGGCCTAAGA	16620
AGGGCGACCA ACTCTACATC CGTCTTGAAG TGGATAAGAA AGACCGTATC TGGGGCCTCT	16680
TGGCTTATCA AGAAGACTTC CAACGTCTTG CTCGTCCTGC CTACAACAAC ATGCAGAACC	16740
AAAACCTGGCC AGCCATTGTT TACCGTCTCA AGCTGTCAGG AACTTTTGTT TACCTACCAG	16800
AAAATAATAT GCTTGGTTTT ATTCACTCTA GCGAGCGTTA CGCAGAGCCA CGTTTGGGGC	16860
AAGTATTAGA TGC CGCGCGTT ATTGGTTTCC GTGAAGTGGG CCGCACTCTG AACCTCTCCC	16920
TCAAACCACG CTCCTTTGAA ATGTTGGAAA ACGATGCTCA GATGATTTTG ACTTATTTGG	16980
AAAGCAATGG CGGTTTCATG ACCTTAAATG ACAAGTCATC TCCAGACGAC ATCAAGGCAA	17040
CCTTTGGCAT TTCTAAAGGT CAGTTCAAGA AAGCTTTAGG TGGTCTTATG AAGGCTGGTA	17100
AAATCAAGCA GGACCAAGTT GGGACAGAGT TGATTTAGGG AGGCTTATGA GAAAATCATT	17160
TTACACTTGG CTCATGACCG AGCGCAATCC TAAAAGTAAC AGTCCCAAAG CAATTTTGGC	17220
AGACCTCGCT TTTGAAGAGT CAGCCTTTCC AAAACACACA GATGATTTTG ATGAGGTCAG	17280
TCGCTTTTGG GAGGAGCATG CCAGTTTCTC TTTTAACCTA GGAGATTTTG ACAGCATTTG	17340
GCAGGAATAT CTAGAACACT AGCATTATAT CATTTGGGTT GGGCTAGTAA TTTCTCCATC	17400
CCTCTGCTAT AATAAAAAGA AATAAAAAGG TTAGAGAGGT TCTTTATTTG AAGGAACATT	17460
CAATAGACAT TCAACTGAGT CATCCAGATG ACCTGTTTCA TCTTTTGGT TCCAATGAAC	17520
GCCATCTTCG TTTGATGGAA GAAGAGCTTG ATGTTGTGAT TCATGCTCGT ACGGAGATTG	17580
TCCAGGTTTT GGGAGAAGAG TCTGCCTGTG AGGAAGCCCG TCAAGTTATT CAGGCTTTGA	17640
TGGTCTTGGT AAATCGTGGG ATGACCGTTG GTACGCCAGA TGTAGTCACT GCGATTAGCA	17700
TGGTCAAAAA TGATGAAATT GACAAGTTTG TCGCCCTTTA CGAAGAAGAA ATTATCAAGG	17760
ATAATACTGG GAAACCTATC CGTGTCAAAA CCCTAGGGCA AAAGCTTTAT GTGGACAGTG	17820
TCAAACAGCA TGATGTGACC TTTGGAATTG GGCCAGCAGG TACAGGGAAG ACCTTCCTTG	17880
CAGTGACCTT GGCAGTGAAT GCCCTTAAAC GTGGGCAAGT CAAGCGAATT ATCCTAACTC	17940
GTCCAGCGGT GGAAGCGGGA GAGAGTCTTG GATTCTTCC GGGTGATCTT AAGGAGAAGG	18000
TGGATCCTTA CCTTCGTCCT GTTTACGATG CCTTGATCA AATTCTTGGG AAAGACCAAA	18060

284

CGACTCGTCT CATGGAGCGT GAAATTATCG AAATTGCGCC CCTTGCCTAT ATGCGTGGCC	18120
GGACCTTGGA TGATGCCTTT GTCATTCTCG ATGAGGCGCA AAACACGACC ATCATGCAGA	18180
TGAAGATGTT CTTGACGCGT TTAGGTTTTC ATTCTAAGAT GATTGTCAAT GGAGATATTA	18240
GTCAGATTGA CTGCCACGT AATGTCAAGT CCGGTTTGAT TGATGCTCAA GAGAAACTCA	18300
AGAACATCCA TCAGATTGAC TTTGTTTCAAT TTTCAGCCAA GGATGTGGTT CGCCATCCTG	18360
TTGTCGCTCA GATTATCCGA GCCTATGAAT ATTCTACTGA AGTTGCACAC GACTGATTTT	18420
GAGGAAGTTC GCCTGCAAAA GAATAGACTT GTTCGGTAAC TGTAAGAAAGT GTTATACTAT	18480
TTTTATGGAA ACAGTATACG ACAAGCACA AAAACTTAAC TCAAAAACT TCAAACTATT	18540
GATTGGTGTC AAAAAGGAAA CCTTCAACT CATGCTAGAA CACCTGAATT CAGCCTATCA	18600
GATTCAGCAC CGAAAAGGTG GACGTCCACG TAGTCTGCCC ATGGAAGACC AGCTCATTAT	18660
GACCCTCCGT TACTTGCGAT ATTATCCAC TCAGCGTCTG CTGGCCTTTG ATTTTGCGGT	18720
CGGTGTAGCT ACGGTAAATG CCATCATCAC TTGGGTGGAG GATACACTTC GTGCGTCAGG	18780
TAGCTTTGAT TTGGACCATT TAGAAGCCCC GAGTGTCTGT GTGGCTATTG ACGTGACCGA	18840
AAGTCCGATT CAGCGTCCAA ACAAACCAA AGCAAAATTT ATTCTGGTAA AAAGAAACGA	18900
CACACCTTAA AAACCTCAAT TATGCTGGAT TTGACGACAC ATAAAGTCTG TCAATGGCC	18960
TTTTCTGACG GACATACGCA TGATTTTACT CTCTTCAAAG AAAGTATTGG ACAAAGTTTG	19020
CCTGAAACGA CGCTTGCTTT TGTGACCTA GGTATTATTAG GCATCTTGAA ATTTTCATGAG	19080
AATACTTTCA TTCCTGCTAA AAATTCCAAA AATCGCCGCC TGAGTGAGGA TGATAAGCAG	19140
TTAAATAAAG AGATGTCAGC GATACGAATT GAAATTGAAC ATTTTAACGC TAAATTCAAG	19200
ACCTTCCAAA TCATGTCAGT CCCTTATCGT AACCGCAGAA AACGTTTCGA GTTACGGGCG	19260
GAATTAATTT GTGCCATCAT CAATTATGAA GTGAAGTAGA TTCCGAACAA GTCTAATATA	19320
CTTTTGAGAG AGGAAAATCC AGTTGTATAG GCTAAAGGTT TTATCCAAAG GTCTGAGACA	19380
ACGATTAGGC ACGATGGAAA GAACTTTTAT GTGGCTGATG ACGATCAGTG CATCTTCCTG	19440
TGTCATAATC ACAGGGCACA AGAAAGTAGG AATTTGAAAA GATGATTGAC CAACTATCTA	19500
AGTATTACAG TTGTAGGATA CTAAGTAAA AGGATATTCC AAGTATTTTA TCTTTATATG	19560
AAAGTAATCC TCTGTATTTT CAGCATTGTC CACCAGAGCC AAATTTTGCA ACTGTAAAAG	19620
AGGACATGCT TTGTCTACCT GAAGGTAAAG CTAAGGCTGA TAAGTTTTTT GTTGATTTT	19680
GGAATGGATC TGACCTGTG GCTGTTATGG ATTTTGTCTA TGCATATCCT GATGAGGAGA	19740
CTGTTTTTAT TGGTTTGTAT ATGGTTGATC AAGCCTATCA GAGAAAAGGG ATTGGTAGTC	19800
ATATTGTGAC AGAAGCACTA GCTTATTTTG CTAAGAACTT TCGAAAGGCA CGTTTGGCTT	19860

285

ATGTTAAGGG	AAATCCGCAA	TCTCAGCATT	TTTGGGAAAA	GCAGGGCTTT	AAATCAATTG	19920
GATGCGAGGT	TAAGCAAGAA	CTCTATACGG	TTGTTATCGC	TGAACAGAGC	CTAGAAGATT	19980
AGAAATGGCA	TCAAGTAAGA	ACTATTTGGA	ATTTGTTTTG	GAACAATTAT	CAGGATTAGA	20040
TGATGTGACT	TACCGTTCCA	TGATGGGGGA	GTATATCTTT	TACTTCCGCG	GCAAGATTAT	20100
TGGCGGCATT	TATGACGATC	GCTTTTGTAGT	TAAACCCGTG	CAAGCAGTCT	TAGATAAGAT	20160
TGACCAATCT	TCTTTTGAGT	TTCCATACAA	AGGTGCCAAA	GAAATGATTT	GAGTGGAAGA	20220
ACTTGATAAT	AAGATGTTTC	TATAAGACCT	AATTTTAGCT	ATGTATAACC	AACTGCCAAC	20280
GCCCAAACCT	AAAAAGAAAA	AGCAAGGGTG	AACGAAGTAA	AAAAGAAGTC	TGCTAAGGCC	20340
CTGTCTTTGC	ACGGGTAAAA	TTTTATATAT	AAAAAGAAGC	TGGGACTAAA	GAGCTCAGCT	20400
TCCTTTGGTT	TATATAATTG	TCATTACAAG	ACGAAAGTGGT	TGGGCGAAAC	TCTGTGACT	20460
TTATTCAATT	TAGAGTTTCT	TATGCACAAT	TGAGTCTGGA	ACGAAAGTCT	CCAGTTGCAA	20520
AGTATACAGT	ACAATAAACC	AACGATGTAA	TAGCTGATGA	CACAAAGCAC	AGTGGGTAGG	20580
ACTTGCGAAG	TCACCCTTTT	CTTTTCAAAA	TTTATACTAA	ATCATTGATA	TCAGTGTACT	20640
CACGATTAAG	TCCTTGAGCA	ACTGGTAGGT	TAGTCAAGTA	ACCTTGATAA	GTAGTCACAC	20700
CTTGACGCAA	GCCTTCATCT	TCAGAGATTG	CTTGTGCGAA	TCCTTTGCCA	GCCAAAGCTT	20760
CGATATAAGG	AAGAGTGACA	TTGGTTAGGG	CGATGGTTGA	AGTGCAGACA	ACCGCACCAG	20820
GGATATTGGC	AACGGCATAG	TGGAGAACAC	CGTGTTTTTC	ATAGACGGGT	TCATCGTGCG	20880
TTGTCACACG	GTCAGCTGTT	TCGATAACGC	CACCTTGGTC	AACAGCAACG	TCAACGATAC	20940
AGAGCCTGGA	CGCATTTGTT	TGACCATCTC	ATCTGTCACC	AATFCCGGTG	CTTTTGCACC	21000
AGGGATGAGA	ATGGCTCCAA	TCACCACATC	AGCATCTCTC	ACACTTGCTT	CAATGTTGAA	21060
TGAATTAGAC	ATAAGAGTTT	GAATTTGACT	TCCAAAGACT	TCTTCTAGAA	CTGAGAGACG	21120
CTTGGAACCTA	ATATCTAAAA	TAGTCACTTG	AGCACCAAGA	CCAAGGGCGA	TGCGGGCAGC	21180
ATGTGTACCG	ACGACACCAC	CACCGATGAT	AGTTACTTTT	CCTTTTGGA	CACCTGGTAC	21240
ACCACCAAGT	AGAACACCAG	AGCCACCAGC	TTGCTTAGTA	AGGAAGTGAG	CTCCGATTTG	21300
AACAGCCATA	CGACCTGCAA	CCTCACTCAT	AGGAACGAGG	AGCGGTAGTT	GTCCTTGATT	21360
GTCACGAACA	GTTTCAGTTG	TTTTTGCTGT	TAACATAGCA	TCTGCTAATT	CTGGAGCAGC	21420
GGCCATGTGC	AAGTAGGTGA	AGAGAAGAAG	ATCGTCGCGC	AAGTAACCGT	ATTCAGAACT	21480
TAAAGATTCT	TTTACTTTCA	CAACCAACTC	TGCTGCCCAA	GCTTCACCAG	CAGTAGCGAC	21540
AATCTCAGCT	CCTTGCTTTT	GATAGTCAGC	ATCAGTAAAG	CCAGAACCGA	GACCAGCATT	21600

286

TGTTTCGATA	AGGACACGAT	GACCACGACT	AACTAAGCTA	TGAACACCTG	CAGGTGTGAG	21660
GGCGACACGG	TTTTCGTTAT	TTTTAATTTT	TTTTGGGATT	CCGATTAACA	TTGAGATAAC	21720
CTACCTTTCA	ATTGACGGTC	TGTTTGTGGT	TGTCACATTC	CAGTTCATAA	ATCAAAAATG	21780
TGACGGTTTC	ATTGTATATG	AAACCGCTTC	AAAAATCAAG	AAAAACTTGT	CATCCAAATT	21840
TTTTTATGCT	AGACTAGTGA	AAATCAAGCT	CTAATGGAGG	GAAAAGTATG	GAATCAATAT	21900
TTGTGAAATT	TGCCCAGTAT	CCGCTCTATG	AAACGGAGCG	TTTATTGCTC	AGACCTGTAA	21960
CTTTGGATGA	TGCGGAACAA	TGTTTGACTA	TGCCTCGGAC	AAGGGTAATA	CACGTTACAC	22020
TTTTCCAACC	AATCAAAGCT	TGGAAGAAAC	CAAGAATAAC	ATTGCTCAGT	TCTACTTGGC	22080
TAATCCCTTG	GGACGTTGGG	GAATAGAACT	AAAAAGCAAT	GGTCAGTTTA	TTGGAACCAT	22140
TGACTTGCAC	AAGATTGATT	CTGTTCTTAA	GAAGGCAGCT	ATTGGCTACA	TTATCAATAA	22200
AAAGTATTGG	AATCAAGGAT	TAACGACAGA	AGCCAATCGT	GCTGTGATTG	AGCTAGCTTT	22260
TGAGAAGATA	GGGATGAATA	AGTTGACTGC	CCTTCACGAT	AAGGCTAATC	CCGCGTCAGG	22320
AAAGGTCATG	GAGAAATCAG	GCATGCGTTT	TTCCCATGCA	GAACCATATG	CTGTGTATGA	22380
CCAGCATGAA	AAAGGCCGAA	TCGTGACAAG	AGTTCATTAT	GTCTTGACCA	AGGAAGACTA	22440
TTTTGCAAAT	AAATAAGCAG	TTGAAAAGAA	ATTTTTCGAC	TGTTTTTTCT	TCCTCTTACG	22500
AATAATCTAA	GAGAGGAGAA	AATATGGAAG	CAATTATCGA	GAAAATCAAA	GAGTATAAAA	22560
TCATCGTCAT	CTGTACTGGT	CTGGGCTTGC	TTGTAGGAGG	ATTTTTCCTG	CTAAAACCAG	22620
CTCCACAAAC	ACCTGTCAAA	GAGACGAATT	TGCAGGCTGA	AGTTGCAGCT	GTTTCCAAGG	22680
ACTCATCGAC	CGAAAAGGAA	GTGAAGAAGG	AAGAAAAGGA	AGAACCCCTT	GAACAAGATC	22740
TAATCACAGT	AGATGTCAAA	GGTGCTGTCA	AATCGCCAGG	GATTATATGAC	TTGCCTGTAG	22800
GTAGTCGAGT	CAATGATGCT	GTTCAGAAGG	CTGGTGGCTT	GACAGAGCAA	GCAGACAGCA	22860
AGTCGCTCAA	TCTAGCTCAG	AAAGTTAGTG	ATGAGGCTCT	GGTTTACGTT	CCTACTAAGG	22920
GAGAAGAAGC	AGTTAGTCAA	CAGACTGGTT	CGGGGACAGC	TTCTTCAACA	AGCAAGGAAA	22980
AGAAGGTCAA	TCTCAACAAG	GCCAGTCTGG	AAGAACTCAA	GCAGGTCAAG	GGACTGGGAG	23040
GAAAACGAGC	TCAGGACATT	ATTGACCATC	GTGAGGCAAA	TGGCAAGTTC	AAGTCAGTAG	23100
ACGAGCTCAA	GAAGGTCTCT	GGCATTGGTG	GCAAAAACAT	AGAAAAGCTT	AAAGACTATG	23160
TTACAGTGGA	TTAAGAATTT	CTCTATTCCC	CTAATTTACC	TGAGTTTCTT	ATTACTTTGG	23220
CTTTATTACG	CTATTTCTC	AGCATCTTAT	CTTGCTTTGT	TGGGCTTTGT	TTTTCTGCTA	23280
GTCTGTCTCT	TTATCCAATT	TCCGTGGAAA	TCTGCTGGTA	AAGTTCTAAT	AATTGCGGGA	23340
ATCTTTGGAT	TTTGGTTTGT	TTTTCAAAAT	TGGCAACAGA	GTCAAGCGAG	TCAAAATCTG	23400

GCGGATTCTG TTGAAAGGGT ACGGATTTTG CCTGATACTA TTAAGGTTAA TGGTGATAGT	23460
CTATCCTTTC GTGGCAAGTC TAACGGTCGT GCTTCCAAAG TCTATTATAA ACTCCAGTCC	23520
GAGGAGGAGA AAGAAGCCTT TCAAGCTTTA ACTGACCTGC ATGAGATAGG ACTAGAAGGG	23580
AAGCTTTCGG AGCCAGAAGG GCAGAGAAAT TTTGGTGGCT TTAATTACCA AGCCTATCTG	23640
AAGACTCAGG GAATTTACCA GACTCTCAAT ATCAAACAA TCCAGTCACT TCAAAAGATT	23700
GGCAGTTGGG ATATAGGAGA AAACCTTGTC AGTTTACGTC GAAAGGCTGT GGTTTGGATT	23760
AAGACGCACT TTCCAGACCC TATGGGCAAT TACATGACAG GACTCTTGCT GGGACATCTG	23820
GACACCGACT TTGAGGAGAT GAATGAGCTT TATTCCAGTC TAGGAATTAT CCACCTCTTT	23880
GCCCTATCTG GCATGCAGGT AGGTTTTTTC ATGAATGGAT TTAAGAACT TCTCTTGCGA	23940
TTGGGCTTGA CCAAGAAAA GTTGAAATGG CTGACTTATC CCTTTCCCT TATCTATGCG	24000
GGACTAACTG GATTTTCAGC ATCGGTTATT CGCAGTCTCT TGCAAAAGCT ACTGGCTCAA	24060
CATGGGGTTA AGGGCTTGGA TAATTTTGCC TTGACGGTGC TTGTCCTCTT TATTGTCATG	24120
CCAAACCTTT TCTTGACAGC AGGAGGAGTC TTGTCCTGCG CTTATGCTTT TATCCTGACC	24180
ATGACCAGCA AAGAAGGGGA GGGGCTCAAG GCTGTACTA GTGAAAGTCT AGTCATCTCC	24240
TTGGGCATAT TGCCCATCTC ATCCTTCTAT TTTGCGGAAT TTCAACCTTG GTCTATCCTT	24300
TTGACCTTTG TCTTTTCCTT TCTTTTGAC TTGGTCTTCT TACCGCTCTT GTCTATCTTA	24360
TTTGTCTTTT CCTTTCTCTA TCCAGTCATT CAGCTGAACT TTATCTTTGA ATGGTTAGAG	24420
GGCATTATTC GCTTGGTCTC GCAGGTGGCA AGGAGACCAC TTGTCTTTGG TCAACCCAAC	24480
GCATGGCTTT TAATCTTATT GTTAATTTCC TTGGCTTTGG TCTATGATTT GAGGAAAAAC	24540
ATTAAAGGAT TAACAGTATT GAGTTTATTG ATTACAGGTC TCTTTTCCTT TACCAAGTAT	24600
CCACTGAAA ATGAAATCAC CATGCTGGAT GTGGGGCAAG GAGAAAGTAT TTTCTACGGG	24660
ATGTAACCTG GAAAACCATT CTCATAGATG TAGGTGGTAA GGCAGAATCT TATAAGAAAA	24720
TCAAAAAATG GCAAGAAAAG ATGACGACCA GCAATGCCCA GCGAACCTTG ATTCCCTATC	24780
TCAAAAGTCG AGGAGTAGCT AAGATTGACC AGCTAATTTT GACTAACACG GACAAGGAGC	24840
ATGTTGGAGA TTTGTCAGAG ATGACCAAGG CTTTCCATGT AGGGGAGATT CTAGTATCAA	24900
AAGACAGTCT GAAACAGAAG GAATTTGTGG CAGAACTACA GGCGACTCAA ACAAAGGTGC	24960
GTAGTATGAT AGTAGGGGAG AACTTGCCCA TTTTGGGAAG TCAGTTAGAA GTTCTATCTC	25020
CAAGGAAAAT GGGAGATGGA GGACACGATG ATACCCTAGT TCTGTATGGG AAATTCCTGG	25080
ATAAGCAATT TCTCTTCACG GGAAATTTGG AGGAGAAAGG AGAGAAGGAC TTGCTGAAGC	25140

288

ACTATCCAGA	CTTGAAAGTA	AATGTTTTGA	AAGCTAGCCA	ACATGGCAAT	AAAAATCAT	25200
CAAGTCCAGC	CTTTCTAGAA	AAACTCAAAC	CAGAGCTTAC	TCTTATCTCA	GTTGGAAAAGA	25260
GCAATCGAAT	GAAACTCCCC	CATCAGGAAA	CATTGACACG	ACTGGAAGGT	ATCAATAGCA	25320
AAGTTTATCG	AACTGACCAG	CAAGGAGCTA	TACGTTTAA	GGGTTGGAT	AGTTGGAAAA	25380
TCGAAAGTGT	TCGATAGGAA	GGATAAATGT	TGTAGATTAG	TGAAATAAAC	TAAAAATTG	25440
TTGCATAATA	ATGATAAAAA	TGGTATAATG	AAAACGTATT	CAATATTGAG	GATATAAAAT	25500
CATTAAAAAT	CAGCAAAAGT	TGTTTTATTA	GTTAGTTTAT	AATCTATTGG	TCTTCTTCAG	25560
TCCAGTGTAT	CTGCTGTGAC	AGTCACTAAA	AGTTACAAGT	ATGATTGGAA	TACGGTTTGG	25620
GAATATAGTA	CCAACTATCA	CGACCATCAG	TATGCTTGA	TTCCGTCATG	GTCTCGTTAT	25680
GACAGCTATT	CTGAGTATAA	AGTTGGCGGA	GGCTGGAAC	ACGCTCGTTA	TGAGGTCATA	25740
AACTATTACA	GCGGAGGCTA	TTAATTCTTA	AAGAGTGAGA	AAAAGGAGGG	CTAGATATGT	25800
TGCAGCTTAC	TCATGTGACC	TTAAAAACGC	GACAAGTCAT	CTTGCAAGAT	GTGGATTTC	25860
CCTTTAAAAA	GGGTAGGGTT	TATGGTCTTC	TTGCTATCAA	TGGCTCTGGA	AAGACGACCC	25920
TGTTCCGTGC	CATTAGCAAT	TTAATTCCTA	TAAGTAGTGG	AAATATCGCA	GCCCCCTCTT	25980
CTTTATTTTA	TTATGAGAGT	ATTGAATGGC	TGGATGGAAA	CTTAAGTGGG	ATGGACTACC	26040
TTCTGCTTAT	CAAAAACATC	TGGAAGTCAG	GTCTGAACTT	GAGGGATGAA	ATCGCCTATT	26100
GGGAAATGTC	TGACTATATC	AGTCTTCCCA	TTGCAAGTA	TTCTTAGGC	ATGAAGCAAC	26160
GCTTGGTGAT	TGCCATGTAT	TTCTCAGTC	AGGCCAAATG	CTGGCTCATG	GATGAGATTA	26220
CAAATGGCTT	AGATGAGTAT	TATCGACAGA	AGTTTTTTGA	TAGGCTAGCA	CAAATCGATA	26280
GACAAGAACA	GCTGGTTCTT	TTAAGTTCCC	ACTATAAGGA	AGAGTTGGTT	GATGTCTGCG	26340
ATAGAGTAGT	AACCATTTCAT	CAGGGGCAGA	TAGAAGAGGT	TTAGTTTATG	AAAGATGTTA	26400
GTCTATTTTT	ATTGAAAAAA	GTTTTCAAAA	GCCGCTTAAA	CTGGATTGTC	TTAGCTTTAT	26460
TTGTATCTGT	ACTCGGTGTT	ACCTTTTATT	TAAATAGTCA	GACTGCAAAC	TCACACAGCT	26520
TGGAGAGCAG	GTTGGAAAAGT	CGCATTCGAG	CCAACGAGAG	GGCTATCAAT	GAAAAATGAAG	26580
AGAAACTCTC	CCAAATGTCT	GATACCAGCT	CGGAGGAATA	CCAGTTTGCT	AAAAATAATT	26640
TAGACGTGCA	AAAAAATCTT	TTGACGCGAA	AGACAGAAAT	TCTGACTTTA	TTAAAAGAAG	26700
GGCGCTGGAA	AGAAGCCTAC	TATTTGCAGT	GGCAAGATGA	AGAGAAGAAT	TATGAATTTG	26760
TATCAAATGA	CCCGACTGCT	AGCCCTGGCT	TAAAAATGGG	GGTTGACCGC	GAACGGAAGA	26820
TTTACCAAGC	CCTGTATCCC	TTGAACATAA	AAGCACATAC	TTTGGAGTTT	CCGACCCACG	26880
GGATTGATCA	GATTGTCTGG	ATTTTAGAGG	TTATCATCCC	AAGTTTGTTT	GTGGTTGCTA	26940

289

TTATTTTAT GCTAACACAA CTATTTGCAG AAAGATATCA AAATCATCTG GACACAGCTC	27000
ACTTATATCC TGTTCAAAA GTGACATTG CAATATCCTC TCTTGGAGTT GGAGTGGGAT	27060
ATGTAAGTGT GCTGTTTATC GGAATCTGTG GCTTTCTTT TCTACTGGA AGTCTGATAA	27120
GTGGTTTGG ACAGTTAGAT TATCCCTACC CAATTTATAG CTTAGTGAAT CAAGAAGTAA	27180
CTATTGGGAA AATACAAGAT GTATTATTC CTGGCTTGCT CTTAGCTTTC TTAGCCTTTA	27240
TCGTCATTGT GGAAGTTGTG TACTTGATTG CTTACTTTT CAAGCAAAAA ATGCCTGTCC	27300
TCTTCTTTC ACTCATTGGG ATTGTTGGCT TATTGTTGG TATCCAAACC ATTCAGCCTC	27360
TTCAAAGGAT TGCACATCTG ATTCCCTTTA CTTACTTGGC TTCAGTGGAG ATTTTATCTG	27420
GAAGATTACC TAAGCAGATT GATAATGTCG ATCTAAATTG GAGCATGGGA ATGGTCTTAC	27480
TTCCCTGCCT GATTATCTTT TTGCTATTGG GAATTCCTATT TATTGAAAGA TGGGGAAGTT	27540
CACAGAAAA AGAATTTTTC AATAGATTCT AGCTTTCCTA TAGGTAGGGA AAATAAGTAA	27600
AACTAACAT AGAGAGGGAA TCAACTTGAT TCTCTCTTT TGATTCGAAA ACCAAACCAA	27660
AATACAAACA CAACTTTTC AAAAAATAAC TTTTATCTT GACAAGAGCT AGAAAACTTG	27720
GTATCATATA AAAGTTGAGA AAAGCAGAAG TGAGAGCTTC TCGCCTTGTC ACATTAAGTT	27780
GCCTGGCCCT ACGGATGAAA AGTTTCGAAG AAACGCTATC ATAACGTGCG GGCTTGATA	27840
TTTACAAGTC CGCTATTGTT TTTCTCTAAT AAAACAAAAG AGGTGAAAAC CATAGCAAAG	27900
CAAGACTTAT TCATCAATGA TGAGATTCTG GTACGTGAAG TTCGCTTGAT TGGTCTTGAA	27960
GGAGAACAGC TAGGTATCAA GCCACTCAGT GAAGCGCAAG CTTTGGCTGA TAACGCTAAT	28020
GTTGACCTAG TATTGATTCA ACCCCAAGCC AAACCGCCTG TTGCAAAAAT TATGGACTAC	28080
GGTAAGTTCA AATTGAGTA CCAGAAGAAG CAAAAAGAAC AACGTAAAAA ACAAAGCGTT	28140
GTTACTGTGA AAGAAGTTTC TCTAAGTCCG G	28171

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7147 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CCGCTCAACT TTTGCAATCA AGGCTAAGTA GACAGCAGCA AATTTATAT TGTATAATTT	60
CTGACTCATA CTTCTCTCTT TCTATGTGTA CTAGTATAAA TAAGAAAAAG AAGGCCGTCA	120

290

AGCCTTCTTT TGATTTATTC TTCTGCTTCA TCCTCTGTAA ATTGACTATF GTACAAGTCA	180
GCGTAGAAGC CACCTTGCGC CATCAGTTCC TCATAGTTGC CTTGCTCGAT GATATTTCCA	240
TCITTCATGA CCAAGATCAA GTCTGCATTT CGGATGGTTG ACAAGCGGTG GGCAATGACA	300
AAGGATGTGC GTCTTCCAT CAAACGGTCC ATGGCTTTTT GGATCAATTC CTCTGTCCGT	360
GTGTCAACAG AAGAAGTCGC CTCATCCAAA ATCAAAAGCG GTGCATCCTT AAGAAGGGCA	420
CGAGCAATAG TCAATAGTTG TTTTGTCTT ACAGACAAGG TCACGGTGTC ATCCAAGATG	480
GTATCATAGC CATCTGGCAA GGTCAATAA AAGTGGTGAA TTCCACAGC CTTACTAGCT	540
TCCATCATTC GTTCATCACT AATCCCTATT TGATTATAGA TGAGATTGTC TCGAATAGTT	600
CCTTCAAAGA GCCAGGTATC CTGCAAGACC ATTGAAAAGG CATCATGCAC TTCTGAACGC	660
GTTCATAGCCT TGGTATCCAC ACCATCAATG CGAATACTTC CCTTATCAAT CTCATAGAAT	720
TTCATCAAAA GATTGACAAT GGTGTCTTA CCAGCCCCAG TCGGCCAAC AATGGCAACC	780
TTTGTACCAG CATGAGCTGT CGCAGAGAAG TCATAGTCTT GAACATTGAC ACCGTCCACC	840
AGAATTCTC CTGCTGACAC GTCGTAGAAA CGTGAATCA GATTGACCAG AGTTGATTTA	900
CCAGAACCTG TTGACCCAAT AAAGGCCACT GTTTGACCAG TTTCTGCTTT AAAGCTAACA	960
TGTTCAATAA CTGCCTCCGA ATTTGCCGCA TAGCGGAAGG TCACATCCTT AAACCTGACC	1020
TGACCTTTGA AGTTTTCATC AGTCAGCTGC ACTTGAACAG GGTTTGGAT AGAAGAATGC	1080
AAATCTAAAA CTTGATTAAT CCGCTTAGCA GAGACCATAG TTCGGGGAAG AACGATGAAG	1140
AGTGCTCCCA TGAGAAGGAA GCCCATGACA ACCTACATGG CATAAGACAT GAAAACAATC	1200
ATGTCACTAA AGAGAGGCAG ACGCGCTATC GGAGCAGCGT CGTTAATCAC ATAGGCCCCA	1260
ATCCAGTAAA TCGCCACACT CAAACCACTT GAAATCCCCA TCATGATAGG ATTCAAAATA	1320
GCCATAAGAC GGTGACAAA CAAATTCAAA CGGGTCAATT CATCATTTAC TGCTGCAAAT	1380
TTTTCAATTT GATAATCCTC TGCATTGTAG GCACGAACGA CACGAATACC TGTAAACTC	1440
TCACGAGTGA TACTGTTTCA TTTATCTGTC AGCCCCTGAA TCAAGGACTG TTTTGAAAG	1500
GCTAGCGTCA TCAAAACGGT CGTCATCAGG ACGTTGATAA TCACTGCCAC AAGTACGGCC	1560
CAGAGCCAGT ATTCTGAATG ACCTAAAATC TTCCCAATAG CCCAGATAGC CATAATTGAA	1620
CCACGCGTTA CCACTTGCAA GCCCATAGTA ATCAACATTT GAACTTGAGT AATGTCAATTG	1680
GTAGTACGCG TCAAGAGGCT AGGAATTGAA AATTTCTTAA TCTCTGTCTG CGAGTAATCC	1740
AAAACTCGGT TAAAAATATC ACTTCTCAGC CTAAGTAGTAT AAGAAGCCGC CACTCGGGAT	1800
GCAAAAAATC CAACTGCAAC TACGGACAAG AAGGCAAGAA AGGACATTCC CATCATCATG	1860
CTTGCCGACT GCCACAATC ATCTAAATTA GTTCTTGAC TACCTAGCAA ATCCGTAATT	1920

291

TTCGAGATAT AGGTCGGCAC TTCCAACCTCT AGATAGACCG AAAAGCAAGT AAAGAGAATG	1980
GCTAGTAAAA TCATCCCCCA TTCTTTTCTA CTAATTCCTT TGGCTAATTT CTTTATTCTC	2040
TCCTCCTATT CCCTTGATAT TTGCGCTGTA GTTGACCGAG AACCTTCTCA AAAATCAGTA	2100
ATTCATCTTC ATCAATGTCT TCCATCAACT GCTTGTCTAT GCGTTCAAAA AAAGCCTTAA	2160
CCTGTTGCAT CTGAGAACGT GCTTTGTCCG TCAGACGAAC AAACCTAGCC CGCTTATCAA	2220
CAGGACTCGC CTCCAATTCC ACCAAACCAT TTTGCACTAT ACGCTTAACC AGATTACTAG	2280
CAACAGGCTT GGTAATATG AGTTCCTGCT CGATATCTTT AATCAAGACC AAGTCTTGGT	2340
TTTTCTCGCG ATTATCCAAA AAACGCACAA CCTGACCTTG CGGCCACCC ATAAATTCAA	2400
TGCCGCAACG TTTGGCTTCC TTTTGACCA TCAGGTGAAT TTGATGACCA AAACGCTTAA	2460
AGACTAACAT CGGTTTATCC ATAATCTCCC CCTTCTAAAT AAAAATAGTT CTCTGGAGAA	2520
TAATTAAATT TCTATGAGAA CTATTTTCTT GATTAAAAA ATCCCAAGTG ATTTTCTCAC	2580
TTAGGATCAT GTTCTATAGG TTAAATTAAA ACCCATCTAC GTTCGTATAA ATCTTTTGGA	2640
CGTCTTCGTC GTCTTCAAGA ACGCTGTAAA GTTTTTCAAA GGTTTCAAGG TCTTCGCTG	2700
ACAATTCAC TTCTGACTGA GGAATCATTT CCAATTCAGT CACTTGGAAT TCTTCAATAC	2760
CAGACTCACG GAGGGCAACG ATAGCCTTGT GAAGGTCAGT TGGCGCTGTG TAAACTGTGA	2820
TTGTACCTTC TTGTGCTTCT ACGTCATCCA CATCCACATC CGCTTCGAGC AATTGCTCAA	2880
AGACTGCGTC CGCATCTTCA CCTCCAAATA CAATAACACC TTTGTTGTCA AAGAGGTAAG	2940
AAACAGAACC TGAAGCGCCC ATGTTTCCGC CGTTTTTACC AAAGGCTGCA CGGACATTGG	3000
CTGCTGTACG GTTGACGTTA GAAGTCAAAG TATCCACAAT TAGCATAGAG CCATTTGGCC	3060
CAAAACCTTC GTAACGTCTT TCTGTAAAGG TTTCGTCTGT GTTTCCTTTG GCTTTATCAA	3120
TCGCTTTATC GATAATGTGT TTTGGCACTT GGGCTTGTTC AGCACGGTCG ATAACGAATT	3180
TCAAAGCTGA GTTTGATTCT GGATCTGGAT CACCTTTTTT AGCTGCTACA TAGATTCTA	3240
CACCAAATTT TGCATATACT TTAGAGTTAG CTCCATCTTT AGCCGTTTTT TTTGGCTACGA	3300
TATTGGCCCA TTTACGTCCC ATTAGGAATC TCCTTTTTTC ACATTTTAAT CTTTCTTATT	3360
ATAACACAAG TTTTTTTGAT TTTCAC TAGA GGAAATGGAT TTTATTAGCA AATCAAGCTA	3420
GGATAGCACT TTACCTGCTA AGATGGTCTT GCCTTTCTAT CTTTATCAAC AGGCACTCAT	3480
CCACATTCAA AAAACAAACT AGACCATTAT CTGCAAAATAG AAAGTTTCAG CCAAGTTTGA	3540
CAAAGTCAGC TCAAATTAAT GTTTGAAGTT TGTAGATATA AGCGACAAAA ACAATCATAC	3600
TGCACCTTTT GTTGACAGTC TACTCCAGAC ATATCATAGT TCAAGTAAAT ACTTTGAAAT	3660

292

TCAACAGTTC TTATAGGCGC TATTGTATTC TAAGAAATCA ATAGAAGAGT TTCTAAGCAA	3720
ACCTCTAATA CTCAATAAAA ATCAAAGAGC AACTAGAAA GCTAGCCTCA GGTGCTCAA	3780
AACACTGTTT TGAGGTTGCG GATGGGGCTG ACATGGTTTG AAGAGATTTT CGAAGAGTAT	3840
AATTTACGTG TTCCCAAGAT GGAGAAGTTA GACTAGTACA CTGGCACTTC TAAAACATTG	3900
CTAGCAATTG ATTTGTTCAT ATTTAATTTT ATTTTTCCTA TAAATGGGTA TTAGATATAA	3960
ACAGCAAAAT ATTTCCGATA CGTGTGCTTC TTGAATTTCC AATCATCTAA AACAAGTAAA	4020
GGATAATCAA TCCCCTGTAT ATCAAGGAAT TGGCTACCCT TTTTACTTTT TTACACATTC	4080
TGTTTGATAG ATTCAATTTA ACATCACGAG CATACTCCAA TGGAAATCGC TAGGCAAGAG	4140
ATAAACTTTC AGATATCCGC AGAGAGATCA TCGCCTCTTT TTGTGCGAAG CATTCTCCTC	4200
TCCTAGTCAT TTTCTACCTT ATCTTCTACC TGAGGATAGA GAGTTGTTC CCAAATAGAA	4260
ATCGTCCGCT TACGCACTAG TGGCAAATCG GTTTTTCAT AAACCGTACG CCACCATTC	4320
CAGGCAAGCC CGGTACACTC TCTAATTTTG ACAGAGAGAT TACGAACATT CCCTTTTAAA	4380
GGATACTAG TGGTAAAGTG AGCCGTTAAA TCCTGCCCAT TTCTGTCCCA AGCCTTAGGA	4440
GTCAAGACTT CCTTACCTTG ATGATCATAG GATAATTCAT TCCAAGTAAT ATAATATTGG	4500
GCAACATAGG CACCACTATG ATCCAGCAGT AAATCTCCGT TTCTGTAAGC TGTAACCTTA	4560
GTCTCAACAT AGTCTGTACT ATTTTGAAAG GTCGCAACTA CATTGTCACG TAAAAAGAA	4620
GTTGTATAGG AAATCGGCAA GCCTGGATGA TCTGCTGTAA AGCGACTGCC TTCTGAATC	4680
AAGTCTCTA CCATATCCAC CTGCTGTTT ACAACTCGGG CACCCGAAC TGGGTCGCCC	4740
CCTAAAATAA CCGCCTTCAC TTCTGTATTG TCCAAATCT GTTCCACTC TGTCTGAGGA	4800
GCTACCTTGA CTCCTTTTAT CAAAGCTTCA AAAGCAGCCT CTAATTCATC ACTCTTACTC	4860
GTGGTTTCCA ACTTGAGATA GACTTGGCGC CCATAAGCAA CACTCGAAAT ATAGACCAA	4920
GGACGCTCTG CAGAAATTCC TCTCTGTTT AAATCCTCTA CCGTTACAGT ATCTGAAAC	4980
ACATCTCCTG GATTTTAAAC AGCATCTACG CTGACTGTAT AATAAATCTG CTTAAAATTA	5040
ACAATCTGAA TCTGCTTTT GCCTGAATGG ACAGAGTTAA AATCAATATC AAGAGAATTC	5100
CCTGTCTTTT CAAAGTCAGA ACCAACTTG ACCTTGAGTT GTTCCATGCT GTGAGCCGTG	5160
ATTTTTCAT ACTGCATTCT AGCTGGGACA TTATTGACCT GACCATAATC TTGATGCCAC	5220
TTAGCCAACA AATCGTTTAC CGCTCCGCGA ACACTTGAAT TGCTGGGGTC TTCCACTTGG	5280
AGAAAGCTAT CGCTACTTGC CAAACCAGGC AAATCAATAC TATAAGTCAT CGGAGCACGA	5340
TCGACCGCAA GAAGAGTGGG ATTATTCTCT AACCAAGTCT CATCCACTAC GAGAAGTGCT	5400
CCAGGATAGA GCGACTGTC GTTGGTAGCT GTTACAGAAA TATCACTTGT ATTTGTCGAC	5460

293

AAGCTCCGCT TCTTCTTTC GATAACAACA AACTCATCGG GTAGCTGATT ACCCTCTTTG	5520
ATGAAACGAT TTTCAATACT TTCTCCCTGA TGGGTCAAGA GTTCTTTTT ATCGTAATTC	5580
ATAGCTAGTA TAAAGTCATT TACTGCTTTA TTTGCCATCT TCTACCTCCT AATAAGTTCC	5640
TGGATTGAGT TGCATAAACT CAGACTTGTT CAGCGAAATC AGCCGTGGTT GGACTAAGTA	5700
ATCCAAAAT TCCTCGTACA ATTCTTCTGA GACATTGCGT CGCCGTCTGG CTAAATAAGA	5760
AGTCGGAATG ACCGTATTAT CCAACATAAA TACCTTATCT AAGTCAATCA AGGTGGTCT	5820
TGTAAGGA TTACGAGCTA GATCCGGCTC TTCTATCATA AAGTTCTGA CCAAACGTCT	5880
GGTCAAGAGA GCTGGTTGA AGGTCTGATT TTAAACCAAC TCTTGTMTT TAGTCATGCT	5940
GTTGTCAATA CAGATATACA TATGATTCTT CACAGCCAAA TCGCTACTAA TAGTCGAAA	6000
AGGCAAATA AGAGCTACAA CATCTCCTCT CTTAATCAAG CAAGAGCACC CCTTTTCTC	6060
CTAATGTAAC ATAGACAGGA TTGACCAAGT CTTCTGATTG ACTCAGAATT TCCAAAGTTT	6120
GAGTTGGCG CGCTGTCAAT TTAGTAGCAT CTTGTCTCTT CAATACAAA TGCTGTCTGC	6180
CAATAACCTT GACAATATAA TCCTTCTCCA AAGCTGACTG GTAAATCCAC ATCAGATGTT	6240
GTCTGTCTG AGAACTCAAG AGAGAAGGAT TTTCAAGCCT CCCGATAGTC TGATAAAAAT	6300
CAAAAACAGG AGCTAACTCC TGCCAATCTG ATTGGCTAGT TGCAAGGCT AGAAAAGGG	6360
CTTTGCGAGC TGATACTTCT TGGTTAGCCT TGAGAGTTAC TTCCCTCC AAGTTTTTA	6420
GAAATCGGGA AACTCCAGAA AGCAAATTTT TCTCTAATG CGAGAAATAA AAACCTTTCG	6480
TTCCCAGACA TAAGTCTTTC ATGTCGCTTT CTCTAGCAA TAAGAGCTCA AACATTTGAT	6540
AGTAAAAGAA AAATATCTGG CACTGGGTCG CGCTCATCTT TTCCTTATCG GCTTCTTTT	6600
TTAACCAGAG CAAGGGCGAC AGGTAGCTGG ATTGAGACAT TTCCTCTACC TCCTACTCTT	6660
TTTTAACTGG AGCATCTGCA CTAGCTGCCA CTTCTTTTGA CTGGATACTT TCCCACTGGT	6720
TAATCTCCTC TGAGATAAGA CCTTCGCATG TCTTGACAAA TAGGGCAAAA GCCTTGGTCT	6780
TTCTGCATA TTTCTCCGTT TGGCATTGAT AGAGGAATTT TTCTTTCTCC AGGAGTTGCG	6840
CAGTTTTTTG GTAAGAAATC CAATTTTCCT TTGCATTATA CAAATGATA ATCCCTCAC	6900
ACAGCAAGCC GAGACTGGAT AAGGCAACCG AAATCAAACG GTAGCGATCA CCTGGCATAG	6960
GAATAGCACA AAAGACAGCT ATGAGGAAAC CTGCCACGAT TTCTGTTATT TTTAATACCT	7020
TATAGCGCCT ACGATGTTGA ACGCTTTTCT TTAAAAATG AGCTATCTGT ACGTCTAATC	7080
GCTCTGTCAG GTACATTTCT TCTGGCGTCA TATTCGTAAC TCCTTTCATT TACTTTGATA	7140
ATCAGGG	7147

294

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 755 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CCGCATGGGA TTGGTGCCT TTTGGGCAAT CTCTTTGACC AAACCTGGAAA CATGTTTTAT	60
GCGCCTGCCT TTACTGCCCT TGTCGGCGGT ACGTCTATAT GATCCTAGTC GCAAAAGTTC	120
CGCGCTTTGG AGCCATTACC ACTATCGGCC TTGTCATTGC CCTCTTTTTC TTGGGAACATA	180
AACACGGTGC TGGTTCCTTC CTTCTGGGAA TTATCTGTGG CCTCCTAGCA GATGGAGTAG	240
CTCATTTAGG AAAATACAAG GACAAAACAA AGAACTTCCT TTCTTTCATT ATTTTCGCCT	300
TTAGTACAAC AGGACCAATC TTGCTTATGT GGATTGCGCC CAAAGCCTAT ATGGCTACTC	360
TTCTGGCAAG AGGAAAATCC CAAGAATATA TCGACCGTAT CATGGTCGCT CCAAACCCTG	420
GAACGTGCCT TCTATTTATC GCAAGTATTG TCATCGGAGC CCTAGTGGGT GCCTTGATTG	480
GACAAGCCTT GAGTAAAAAA TTTGCCCAGA AAATCTGATC AGTTAAAAAG AGCCACGCGG	540
CTCTTTTTTA TTTATGGCTC AATTTCTTAG TCAAGAAATC TCCCAAGAAT TGGATTGCAA	600
AGATAATCAA AATGATAATA ATGGTTGCCA AGATGGTCAC ATCGTGATTG TAGCGGTAA	660
ATCCATAAGC GATGGCTACG TTACCGATAC CACCAGCTCC AACCGCACCG GCCATAGCTG	720
TTtccCAACA AGGGaAtCAA GGTcACAGTC GTCAC	755

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3010 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TTCAATGGT ATCTCAATCA ACGGTCTTCA CATGGTTTCA ACTGGTTTGA CTCTTGAAAA	60
AGCGAAAGCT GCTGGTTACA ACGCAACTGA AACAGGCTTT AACGATCTTC AAAAACCAGA	120
ATTCAATGAAA CATGACAACC ATGAAGTAGC AATTAAGATT GTCTTTGACA AAGATAGCCG	180
TGAAATTCCTT GGTGCCCAAA TGGTTTCACA TGATATTGCA ATTAGCATGG GAATCCACAT	240
GTTCCTCACTT GCTATCCAAG AGCATGTGAC AATTGATAAA TTGGCATTGA CAGACCTCTT	300

295

CTTCTTGCCA CACTTCAACA AACCATACAA CTACATCACA ATGGCTGCCC TTACGGCTGA	360
AAATTAAAAA TGAATGAGCT ATCTGGCCTT AAGTTAAGGT CAGATAGTTT TTAGCTAATT	420
TGTCCCCATA CAATTATAGT TTTTATTATCT TGTGCTTCAT TCTGTTCTGA CTTAAAATGA	480
AAAGGTAGCT ACCAATACAA ATGATGAGGA TAAAACAAAT GACTGAAAAT CGTTATGAAC	540
TAAATAAAAA CTTGGCACAG ATGCTCAAGG GTGGTGTAT TATGGATGTG CAGAATCCTG	600
AACAGGCTCG TATCGCAGAA GCTGCTGGTG CGGCAGCTGT GATGGCCTTG GAACGAATTC	660
CGGCTGATAT TCGTGCAGCT GGAGGAGTTT CCCGCATGAG CGACCCAAAG ATGATTAAGG	720
AAATCCAAGA AGCGGTAGT ATTCCAGTAA TGGCTAAGGT CAGAATCGGG CATTTTGTG	780
AAGCTCAGAT TTTAGAGGCT ATTGAAATTG ATTATATCGA CGAGAGTGAA GTTCTATCTC	840
CAGCTGATGA CCGTTCCAT GTGGACAAGA AAGAATTCCA AGTTCCTTTT GTCTGTGGTG	900
CTAAGGATTT GGGTGAAGCC TTGCGTCGTA TCGCTGAAGG TGCTTCCATG ATTCGTACCA	960
AAGGAGAACC AGGACAGGG GATATCGTCC AAGCTGTTCG TCATATGCCG ATGATGAATC	1020
AGGAAATTCG CCGCATTCAA AACTTACGTG AGGACGAGCT TTATGTTGCT GCCAAGGATT	1080
TGCAAGTCCC TGTAAGATTG GTCCAATATG TTCATGAACA TGGAAAATTG CCAGTTGTAA	1140
ATTTCCGCTGC TGGAGGTGTT GCAACGCCAG CAGATGCTGC GTTAATGATG CAATTAGGGG	1200
CAGAGGGGGT CTTTGTCTGGT TCAGGTATTT TCAAGTCAGG AGATCCTGTT AAACGAGCGA	1260
GTGCCATTGT TAAGGCTGTG ACTAACTTCC GTAATCCTCA AATCCTAGCT CAAATCTCTG	1320
AAGATTTAGG AGAAGCCATG GTTGGTATTA ATGAAAATGA AATCCAAATT CTCATGGCTG	1380
AACGAGGAAA ATAGATGAAA ATCGGAATAT TGGCCTTGCA AGGGGCCTTT GCAGAACATG	1440
CAAAAGTGCT AGATCAATTA GGTGTCGAGA GTGTAGAAGT CAGAAATCTA GATGATTTTC	1500
AGCAAGATCA GAGTGAAGT TCGGGTTTGA TTTTGCTTGG TGGTGAGTCT ACAACCATGG	1560
GCAAGCTCTT ACGTGACCAG AACATGCTAC TTCCCATCCG AGAAGCCATT CTATCTGGCT	1620
TACCAGTGTT TGGGACCTGT GCGGGCTTAA TTTTGCTGGC TAAGGAAATC ACTTCTCAGA	1680
AAGAGAGTCA TCTAGGAACT ATGGATATGG TGGTCGAGCG TAATGCTTAT GGGCGCCAAT	1740
TAGGAAGTTT CTACACGGAA GCAGAATGTA AGGGAGTTGG CAAGATTCCA ATGACCTTTA	1800
TCCGTGGTCC GATTATCAGT AGTGTGGTG AGGGGTAGA AATTTTAGCA ACAGTGAACA	1860
ATCAAATTGT TGCAGCCCAA GAAAAAATA TGTGTGTAAG TTCTTTTCAT CCAGAATTGA	1920
CTGATGATGT GCGCTTGAC CAGTACTTTA TCAATATGTG TAAAGAAAAA AGTTGAGATT	1980
GAATTTCTCA ACTTTTCTAC ATGTAATAAA CAATAGCGAT GTATTGAAGT GCGGACGCAG	2040

296

CTAGGATAAA GAGATGCCAA ATCATGTGGA AATAAGGTTT TTTCTTGGCA TAAATCCAG	2100
CTCCAAGTGT ATAACAGAGT CCGCCAGTTA CCATGAGACT CCAGAAAACG GGTGTCGTTT	2160
GACTGATAAT GGCAGGAATG ATAGCCAGAA CCAACCAGCC CATAATCAGG TAAAGAGCAA	2220
GGCTAAATTT CTCATTGACC TTTTGTAGCA AGATTTTATA GAGAATACCA AAGATGGTCG	2280
TTCCCCATG GATGACAATA ATCAGATAGC CAAACCAGTT ATTCATCAAG GTCAAGACAA	2340
CGGGCGTGTA TGAGCCGGCA ATGGCAACGT AAATCATAGA ATGGTCAATG ATTCGCAAAA	2400
CATATTTGTG GGTGCAACCA TAGGCCATAG AGTGATAAAT GGTGGATGAT AGGAACATGA	2460
GAAAGAGACT GATGACGAAA ATGGAAACGC CGATAGAGGA TAAAAATCCG TGTGCTTCAT	2520
AACTATAGAT GGATGAAATA GGCAGCAAGA TAAGCATGAT GACTGCACCC ACAGCATGGG	2580
TCACGCTATT AGCAATCTCC TCTCCAAAAC TGAGTTGTTT GCTGAGTTTA AGACTAGTGT	2640
TCATTGGATT ACCTCCTCTT GAGTATGATC GATTAAGTCT AGAGTTTGAT GATAGAGTTT	2700
AACGGTTTGG CAGCTGGTTT GGATAATAGG GTTAGCTGGG TCAATTCCTT GGTTCATGTA	2760
GTCCACAAAA GCATCGTAGA GTTGGTCTGA ACTTGCTTGA GTTTGTAGAG TATTAAGTGT	2820
CTGGGCTATT TCTTGAATAG AAAATACAGA CTTGAGGGTT GTGATAGCAA TCAAACGGGC	2880
AATCTGTTGG CGTTGGTATT TTTTGTGTC AGGCTTTGTC AGGTAACCAT TTTTCACATA	2940
ATTGTTGACC ATAGATGCTG TTAGGCCCTT GTCTTTATTA GGAGAGATAG GGGCGCAGAC	3000
CTGATTGACA	3010

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15213 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CATAAATCGG TGCAAATAAC TTAATAGTGA AGTAGCCATT TCTTTCGTAT TTACCTGAGG	60
CATATTCCTT AGACGAAAGA ATATTATTAT CAATCAAATC ATTGAATGAA CGTAGTCTTT	120
CAACTTCTTC TACTGTTAGA TTTCTGACAA CATTTGTTGC ATAGACCTTA TTTCCATCAG	180
GATCAGGATG GTACTCATTT GTAACTTTTC TAAGAAGTTG TTGTTTTTGA TTCGTATCCA	240
ATTTAAGAAT TGAATTTTCT TCGAGATATT CCAACATATA AACACGTC AACTGTTGT	300
GGACATATTG CTTCAAATCA TCTGCATTAT TAAATCTTGT AGTTGGATCA AGTACTTGTA	360
ATCGTCGACT TTCTGTACTA TCAGATTTTG AATGTTTCAA GATGGAGTTG ATGGTAATGG	420

297

TCGCATCATC TGGATGGTCT GGTGCTTGTA ATAATCCTTT AGCAAAGAAC TCTGGTCCCA	480
AGCCACTTCT TCGACCATAT CCTCCAAGAT AAATGTCCTG ATCTGAGTCA TGTGTCATCT	540
CATGCGTATA AGTAATAGCT CCATCCTTAT CCAACATTCG ATAACCCATA TAATAAACTG	600
CATCACCTGT AGCATAAGCA CCGTGTTGAT TATGCCCAAC TTTATTTCCA ACAGGTCCAA	660
AGAAATGTTG CATGTCAGGA TTTGGATTAT CAAAATCTGC CACTTCTGTA GCTTTCCTTA	720
CGGTATTATC ATCGCCAAAT TTATAAGCAT CGTAAAGCAA AATATTTCTA TAAAGTTTTT	780
CACGTGCATT GTCGTCTAAA ATACGATACC AATAATCGTA GTGATCTCGC TGACGTTTGG	840
CTGTTTCACG CGCATTTTCT TCAACAAAAT CATTGAGAGC CTTGCCCGCT TTATGGTCAC	900
TACTGCGGTA GCGATCATAA GCTCCAAATC CTAGACTAGA CATGGTCGAG ATGACAAATA	960
CGGATCTCTC TGGCAAGGTC AGGAGAGGCA AGACCATATT GCGGTATTTT CATGTGGCAC	1020
TCGTGATACG ATCATAAACA CCGATAGAAT ACTTGGTGCC AGCTAACCCCT TGCTTCGTTT	1080
TCACCTCTTC GATAGTGGAT TTTTCTTCGA CAATGTAAGC CTTAGTCTCT GATTTAAACC	1140
AGTCATTATP GCTTGTATTT GGTAAAAAGA CTTTTCGGTA ATGTTCCAGC GTGCTAAACA	1200
AATCTGTCGT TCCATGTTGA CTGGCAAGAC TGATACCATA AGTATCGACA TTATTCTTAG	1260
CTAGAAGATT GTTAAAGCCA GATTACCCA ACTCAATCAG AGTATCTAAT GGTGAAGCAT	1320
TCCCCTTACC AAAGAAGTCC AAATGGTACA GAACTAGGTC TTTGACATTC ACCTGACCAT	1380
AGCTAAAGTT ATACCACCGT TCCAGATAGG TCAAGCCAAG TAGCAAGGCT TCCTTGTTGC	1440
GTTTGATTTT ATCTACAAGA TAACCTTCAG TGACGGGGTT AGCACTAGCC AGTCCAGCAT	1500
CCGCTGACAA GAGTTTTTTC AACTGTCTT CCAGTTGTTG TTTTGTTTTG GCGAACTGGT	1560
CTTCTAGATA GAGCTCAGTT TGCTTGACGT TTGGAGAAAT ACCCAGCGTC TTTCTGATGG	1620
CTTCTGAATG ATAGTCAACC TTTTGTAAGT CAGGTAAGAC TTGCTTGATG ATAGAGGTTT	1680
GGTCATACAG GAATTGGTTT GCGGTATAGA GAAGTCCAGT ATTGCCCAGA CTATATTCTG	1740
CTAATTGTC GAAATCATTC TGGTATTTGA GATCCAGCTT CTCAGATAAA TCATCCTTGT	1800
AGTGAAGCAA GAGTTTGTTT GCAGTCTGTT TGTTAGAAAC AATGTCTGTG ATGACTTGGT	1860
TGTCCTTCAT CATGACTGCT GACAAGAGTT CTTTTTGATA TAAAAGACTG TTCTCATTGA	1920
CCAGGTTTCC GTATTTGACG ATGGTTGCCT TGTTGTAGAA AGGTAGCAAT TTTTCAATGT	1980
TTTTATAAGT CAAGTTGCGC TTAGCTTGAT AATAGGCCAC CTTAGAAAAA TCACTGTCTT	2040
TTTTGCCACT TGTTGAAAGT GGCTCCACTG TTGGTAAAT GAGAGGATTG ATTTCTGCTT	2100
TTTTGCTTGC AATTTGAGAA GCATCTAGCA TTGTTCTCTT TTCTTCAAAG GATTCCTTGC	2160

298

TGACGACCTC ATCCTTGACC AAGGTGACAT TGTAGACTCT GTTGGCCTTG CTGCTGAATG	2220
TGTCCTTTAC CTTTCATTCG TTATAGTGGT AACCAGTGAT GGCATTTCCG TTGGTTACAT	2280
TAACATCGCT GAGAACATTG GTCAAACTTC CAGCATGCCT AACATCACCA GAAGTTCGAT	2340
CCCACAAATT GCCTGCCACT CCAGCGACTC TACCAAAGTG CTTGACATTG TTGATATCAC	2400
CTTCAGCATA GCTATCTTGG ATCTGTGCAT CTCGGTCTAC TAGGCCTGCA AGTCCACCCA	2460
CAGTCTGATC TGAAGTATTT GTGTTAGATG AAATGGCTAC TGTGCTTTT GACTTAGTAA	2520
GTAAAGCCTT GTCACCTGTC AAATGACCGA CCATACCACC GATATTGTAG GCAGCAGTCG	2580
TTTCATAAGT GTTGATAATT CTTCCCTTGA AACTGCTCTC TGTGATGCTT GATTGCTCAG	2640
CCTTAGCCAG CAAACCACCG ATACCACGTT CACCAGCCAG AACACCATCG ACGTGAACCT	2700
GCTTAATTTT TGTGTTATTC TGAGCTTCAT TTGCCAGTGA ACCGATATCA TCTTTCCCTG	2760
AAATAGCAAC ATTTTTTAGA CTCAGTTTTT CTAAGTGTAG ACCACTCAAG TTTTCAAACA	2820
GAGGTTTTTT CAAATTATAG ATAGCATAAT TCTTGCCATC TTTTTCACCG ATTAACGAC	2880
CAGTAAAGGT GTCCTTGATA TAGGATCTTT CATCAGGACC AAGCTCCACT TCGTTAGCAT	2940
TCAGGCTGGC CGCTAAATGA TAGGTTCCAG AGGGATTTTG GTTTATAGCT TTGACCAGAT	3000
TACTAAAGGA AGTAAAGTTT GTTGTTCCTT CTGTTCCCTT CTTAGCTAGA TAGAAGGTAA	3060
AATTATCTTT ATATCTGCTT TCTATCTCCT GCTGAAGCTT CTCTACTTTT GCTGTGATTT	3120
TATAAAGGAT TTTATCATTT TTTCTTTCCT CTGATATTGA TGCTACTGGT AGGTATACAT	3180
CTTTGAATGA AGAAGATTC ACTTTAACA AGTAGCTATT TGGATTGCTT GGAACCTGCT	3240
CTAACGAAAT GTGTTGTTTA TAAGTACCAT TTGACAAACT GTATAACTCT AGGTCGGAAA	3300
CATTCTTAA TTCAAGTGT TTCTCTGGTT CTTCTACCTT TTTATCAGGG TCTAGTTCAT	3360
TTTCTTGTTT AATTTCTTCG TTTCCATTG AATTGGATGT GTTTGATTCG GTTGAAACAT	3420
CCTCAGTTGA ATTTCCGTTT GATGGTCTG GTTCTGTTTG TCCATTCTCT GATGTTGTAT	3480
TACCTGAATT TTCTGGTTTT GTTGCAGTTC CGTTTTTTTC TGGTTGATTT GATTCTTCAA	3540
CTGGTGGTTT TGAATCACTA GGTTTATTGG ATACTTCTCC AGTATTTTCG TTAGCTATTT	3600
TCCCAGAGTT TGTTTGTGTT TCTTCTGCAG GTTGAAGTGG TTTTCTGTT TCTTGATTTG	3660
AGGTACCTTC TACTGTGCCT TCATTTGGAT TTAGTGGAAC TTCTTCTACA GTTTTTCTG	3720
AATTTTCATT TTTAGAGTCA TTATGTTCTG GTTTATTTGA TTCTCCAAC GAGGTTGTCTG	3780
AATCACTAGG ATTACTGGAC ACTTCCCCAG TATTTTGTCT AGATGTATCT GGTGATACTT	3840
TCTCTGAATT CGTTGTTGAT TCTTCTGCAG GTTGAAGTGG ATTTTCTGCT TCTTGAATTG	3900
AGGTTCCCTC TGTAGTACCT TCATTTGGAT TTAGTGGTGT TTCTTCTGTT GGTTTTACTG	3960

299

GAACCTCTTC AGTTTTTCTT GGACCTTGTT CTTTGGTCTT CTCAACCGGA GTTTCAGGTT	4020
TTACTTGCTC AATATTACCC TTATATTCTG GAAGCGGTGC TACCTGCTCT GTTTCACCTT	4080
TATCACTTAC CACAGTATCT GCGGACTCTG GTTGAACCTC AGTCTCACCT TTGTCGGTCA	4140
CAACTGCTTC GGGTAATGTA GGTGAACTT CTGGTTCGCC TTTGTCACCT ACTACAGCTT	4200
CGGGCAACTC AGGCTGAATT GCGGGTTCAA CAATAGCTCC AGACTGTACG TCCTTATGTT	4260
CTACACCAGT CTCAGGTTGT TCCTTTATAA CTTGAGTTTT TTTAGTACCT TTTTCGACTA	4320
TTCTTGGAAT AGGCGCAGTC GTTGAAGTTG AAACAATTTT TCGCGAACT TCTTCTTGT	4380
TTACAGAGAA TATTCTGACG ATTTCAACTT TCTTACCTAA TTTACCTTCT TGTTTTACTC	4440
TTACAGTTCC TTCAGCTAAA TCAGGATTTT CTTGAATTTT TTCTTGAAAA TCTATTTTGT	4500
TCTCCATAGT TTCCTCACGA TATAAGAGTT CAGGTTTGTT CAATTGACCT GATAAACTT	4560
CATCCTGTGG ATTTAATGTA TTTACCCAG TCTTTTCTTT TGGAGAAATC TTCTCCTCTT	4620
TCTTCGTTTC TAGATTCTTA TGTTCGGCTA ATTGTTCTTG AGAATCTGAA GATTGTTTCT	4680
CTTCTTTTCT TGGATTGATT AATTCAGTAG AGAAAGGTTT TTCAACTACT TGAACCTCTG	4740
TCGGCTTAGT TGAAGAAACA GGTGTTTGTT CCTGAATAGC TTGTACTGTT GATGGATGGT	4800
CTACAAAATT CGGTGTAACA TTATAATCCA CCTTTTGTTG TTTTGTAGGA GTGGCAACTG	4860
AACTCTTTTG ATTACTTACT TCAGACTCAG AAGTCGTTTT TCCCTCTTTG ATATATCCAA	4920
TATAAGTGTA ACCTGAAATC TCTTTAGGAA GAGGTAATTT TTCTCCAGAG GTCAATTAT	4980
AGTCCGTATT GTAATTTAGC AAAAGATGAT TTTCTAAAGC ATGGACTGAA ACTAAGACAC	5040
CATTTCTTAT CCCTGCAACC AATACTAAAT GTAATACCGT TTTATCTTAA ACCTTTTCTT	5100
TGGAAACAGC AAAAATTAAA ATTCCCATAG CAGCTAAGCT AGCACCAGCA ACTAGGGCTT	5160
GCCTCTCATT CTTGCTTCCA GTATTTGGCA ATTCCGCCAG TTGATTTTGA GAATTTAACT	5220
TATAACAAG ATAATAAGTT TCATCATCAT TCTCCACGTA TGTCGGAATA TCATAGACAA	5280
GCTGCTTCTT TTCTTCTGAT GATAGCTCTG AATCTGCCAC ATATTTATAG TGAACCTCCG	5340
CAGTTTCTTG AGCATCCACA GATGAACTAG CTAATACAGA CATAAAAAAT AAACCTGAAA	5400
TCGTTGCAGA TACAAGTCCT ACTGATAATT TTCTAAATGA AAAACGCTCT TGTTTTTCAC	5460
CAAAATACTT TTCCATTATT CCTCCTTGAA ATAAAATTTA TATATGTTAC AAAGACCTTT	5520
ATTATATTAG TGTATTATCT ATTATCTATA GAAAAGGCAG TATACCTTAA TTATACTCTT	5580
AATTTACAAA AAAGCTTAA AATTGAGATG CGCTTTCATA CTTTGTTTTA TATTATTTGG	5640
AGGTACAATA ACACCTACCA TGAAATTTAC ACGGTAGGTG TTACTCATAT CACTAATCGT	5700

300

TCTAAAAATG GTTTGAGGCA GTTGAGGAGA ATTCCCTTCTA TCCAGCTTCC TTGTGCTGAT	5760
GAGCGATGGT CTTCCTGCAG GCTTTTTTTT AGAAAAATCTC GGACTTGTTT TGGTGCGATT	5820
TCAAATTCAA AGGCTTTCAT TTTATAGAAA AAGTCGATGA GATGATCTGA CAGGTATTCA	5880
GTGAAAAGG GTACTTCACC ACTTTTTCTA TATTCTAATA AGAGTCTAGA AAATCGAGCT	5940
TTTCTTCAG GAAGCTCACG AAAATAGGAA TTGAGGATCC AAGTCTGCTT CTGTTTTCTT	6000
TCAATTGGAT CCTGACTGGC AATTCGTTGG TCTTTTTCCA GCTCTTTTG GTATTGTTG	6060
GCCTTGATAG CTCGTCTGC TCTATTTTA CAAAAAGAA TTTTTTCCA CTGCGTTCT	6120
TCTTGAGTCA GGGTCTCTGT AAAGCCAAAG TAATCTTGAT AAGCACGCTC TCGGGTCCC	6180
ATGGCTAGAA CCAGATTGTC TGCATATTGC TTGGCGATTT TATCCCTCTT CTGCGTTCT	6240
TTCTCTGCCT GGATACGGAG TTCTTGTTCTG TAGTCAATTT TCTCCTGCC TAGCTTGACA	6300
AGGTAGAGTT GGTCAATCCG TTTCCCAAGT AAAAAGGGT TGATACACTT TTCAAGGACT	6360
TCTTCCATCC GAGCCTTTTT CTTTGGTTCC GCCTTGGTCC AACTTCCTCC CTGAAAGACT	6420
TCTAGGAAAA GCTGGTAGTC TCTCTAGGC GCAAATTGAT TGCCACGATT GGGTTTGAAA	6480
ACACCTTTTT CCCAGAGCCA TTTAGAAAGT CGCTCGTCAA AGTTACTTTT ATGACCTTG	6540
ATTTTTTCCT TTTTCTGAGC TTTTCTGGTT AGATTTTCAA CCTTCTGAG CAGTTTTTCT	6600
TCCTCTTCCA ATTGCTGGTC AAGGGACAAT CGATGAAAAT GACGAACACA GTCGCTACCA	6660
ATTGGAAGA GCGTTGGCC TGTGACACCG TTAAAGAGTT CATAAGCGTA TTTGATGGCA	6720
TTTCCACAGA CACAATTGCT ACGGCCGATA CCGTTAAAA TAAAGGAAAC TTCATTCCAT	6780
TCCTTGGTAG CTTGTTCCCA AGTATCCGCT TTGGAAGCCT GTAAAACTGC ATCGTGCAGG	6840
GATTTTCTAA CTGGAAGTGT CATGAGGTCT CTTTCTAAT ACTCAATAAA AATCAAAGAG	6900
CAAACTAGAA AGCTAGCCGC AATCAGCTCA AAACACTGTT TTGAGGTTGT AGATAGAACT	6960
GACGAAGTCA GCTCAAAACA CTGTTTGAG GTTGTGGATA GAACTGACGA AGTCAGTAAC	7020
CATATATACA GCAAGGCGAA GCTGACGTGG TTTGAAGAGA TTTTCAAAGA GTATAAGTTA	7080
TACTTTTACA ACTTGAACCT CGTCTTTACC GAGTAAAATC AAGTATTTT CAATATTTT	7140
AATCGAATAG GCTCGTGATA AAGCCTCTTC GTATAGAGCT AACTGACCAC GATAGCGGTC	7200
TACGAGTTGA CTTGTTTCAT CATAGCGGTC TGTCTGTAG TCGAACAGAA CAATTTTGT	7260
TTCTGTAAGC AGATAGCCAT CAAGGATACC ACGGACAACA AAGTCTTCCT GACTCTTTG	7320
GTCTCGTTTG AGCATGGAGA AAGGTTGCTC GCGATAAAGA TGGTCGGTAT TAGCAAGAAT	7380
TTCTGACCG AGTACTGTGT CAAAGAAAGC AAGAATTTT TCAAGATTGA TCTGTCTCT	7440
GACAGCTTGG CTAGTTTGAA CTTGTTTGAG TGTCTGTGT AGGCTAGCAA GGGTTAGTTG	7500

301

CTGGCTGAGG TCAATTCTCT GCATGAGTTC GTGAGTAGCA CTACCAATCT CAGCTCCAGT	7560
TACCTTTTCT TTGGTTGAAA AATCTGGCAA ATCGAAGCTG ATTTTCTTGC CTA CTGACTG	7620
ACCTTGACCA GCAATCTCGA CACCTTCCAT ATCCATAACT GGTTCGTAGA ATTTCTTGAT	7680
TTGACTTGGG GTTGAACAC TAGGAAGTTC AATAGCTGCG CGGTGAAGAG TATTATAAAC	7740
TTCCACCTCC TTCAGCATTT CCAGAGCTTC TTTGATGGTA TCTGACTGAC GATTGTCTGC	7800
TTGGGAGCTA TCTTGAGAG GACTCTTGGT TTCCAACCTCT CCGATAGCTT CTCTGGTCAA	7860
CTGATCTTCG CCAATAAAAC GATAACTAAA GTTGAGCTTG TCCTTAGTAA ACACTTTACT	7920
GATAGCCCAA AGCCAATCTT GGAAATTCG TGCTTGCACT CTAGTATTGC TATTTAGTTT	7980
CCCATTTTGG GCTGCTGGGT ATTCCTTGGA TTCCAGCTTT TCACGAGAAC CCTTGCCGAC	8040
AAGATAGAGC TTTTCTCAG CCCGCGTCAT AGCAACATAC AGCAAACGCA TCTGCTCAGA	8100
ATAGCTTGCT AGCTGTAATT CCTCTTCGTT CTGCCTATAG GTCAGACTAG GAATGGAGAG	8160
TTTGATGGTT TTAGGATAGT GGTCTTCTAC TGCCCCGTG TCCATCTTGG CAATATATTT	8220
GACACCAAGA CCATTCTGAC GACTGAGAAT GACTTCTGAC ATAGAGTCTT GCTTGTGAA	8280
ATCTTGATCC ATATTGAGGA TAAAGACGTA AGGAAACTCC AGCCCTTTAC TCTTGTGGAT	8340
GGTCATGAGC TCTACTGCAT CTTTGGCGG TGCGACGGCC ACGCTTGCCA AATCGTGCTG	8400
GGCTTCTAAG ACTTGGTCAA TCATACGAAT AAAACGCGAC AAACCTTTGA AATTGCTCTT	8460
TTCAAATTGA TCAGCACGCA GTGCTAGGGC ATAGAGATTG GCCTGCCTAG CAGGACCATT	8520
CGGCAAAGCC CCAACATAGT CATAATAAAA ACGGTCGTTG TAAATCTTCC AAATCAAGTC	8580
ATAGAGAGAG TGGGTTTGG CATAACAAGC CCAAGAAGCT AGGATATCCA TGAATTGCTT	8640
TAGTTTTTCA GCTAGAGCTG TGTGAATCAA GCCTTTTGA CTACTTGCCA TTTTGTGTC	8700
ATTGACCACT TTCTCATAGA GATTTTCGTG GATTTTATCC TCTGCTTCT GAAGGGACAA	8760
ACGTGCTAGC TCATCCTCAT CAAAACCAA CATTGGAGAC TTCATAAGGG CAACCAAGGC	8820
GTAGTCTTGC AGGGGATTGT GAATGACACG AAGAGTGTCT AGCATGACTT GCACTTCTAG	8880
GGATTGGAGA TAATTGTTTT GCTCTCCGTC AGTTTGGACA GGAATTCCGT ACTCAGACAG	8940
GGCGAGGAGA ATCTGGTCAT TACGACTGCG GCTGGAGGTC AGAAGGGCAA TTTCTTAAA	9000
GGCAACACCT TTTTCTTGAT GAAGTTTCAG AATCTCCTTG ATAAC TAAGC GCATTTGCCC	9060
TGTTAGTTTC GTTCTGTTT GACTCTCTTC TTCCTCACCT GTATCGTCCT TGTGCTAGAG	9120
GAGAAATGCT GCCTGTGTT CTGGATTGGG AGTCAGTTTG GTATTGGCAA AAACAAGCTG	9180
GTGCTTGTTA TCATAGTTGA TTTCGCCGAC CTCTGGTCC ATGAGACGTT CAAAGACATC	9240

302

ATTGGTTGCT	GACAGCACTT	CTGAACTACT	ACGGAAATTT	TCCTTGAGGA	TAATGAGCCT	9300
GCCTTCTTGG	GGATTTTGCG	CATAGCGTTG	GAATTTCTCA	TTGAAAATCT	GCGGGTCTGC	9360
CTGACGGAAA	CGATAGATGG	ATTGCTTGAT	ATCTCCCAACC	ATAAAGCGAT	TGTGGCCATT	9420
AGACAACAAT	TCCAGCATCC	GTTCTTGAAT	ATGGTTGGTA	TCCTGATACT	CATCGACCAT	9480
GACTTCATGG	AAGCGCTCCT	GATAAGACTC	ACGAACTTGT	GGGAAATTTCT	CTAAAATCTC	9540
AATGGTGTAA	TGGCTGATAT	CAGCGAATTC	GAAGGCATTT	TCCTGTCGTT	TTCTCTGACG	9600
ATAAGCCTCT	ACAAAATCGC	TCATGAAAGA	TTGGAAGGTT	TTAGCTAGTT	TCCAAGTGTC	9660
TCCATGATAA	CGTTCTTGAT	AGTCGAGAAT	CGCTATCTGG	TCTGATAATT	GTCCTAGTTT	9720
AGCAAACTGG	GTCTTTCTCT	CTTCGTTGTA	GGCATCAGCC	AGGGGCTTCA	AATCAGCCTA	9780
CGGCTGGCAT	TAGTCAGAGC	TCGACCGTTT	TTCTCCTTAG	AGATGGCGAC	AACACGCGCA	9840
AGCACTGCCT	GATAAGCCTG	ACTATCGGAC	TCCTGATTTA	GGGAGCCAAT	TTCATCCAGA	9900
ATTAAGTAA	CATTTTCTAA	ATAGGCAGCC	TTTGCAAAT	CCTTGGCATC	GTTATCCAGA	9960
TGGTAACGGA	AAAAGCTTTC	CAAATCCCAA	AGGGCTTGTT	TGATTTGCTC	GGTCAGTTT	10020
TCTTTTTCAC	TGGTAAAATC	AGCTTTCTCA	AATCCTTTGA	GGAAAGATTC	ACTCAGCCAC	10080
TTTGTAGGAT	TACTGGTGGA	TTGGAGGAAG	TCATAGATTT	TATAGACCTG	CTGGCGCAGA	10140
CCCCGTTCTG	CCTTGCCACG	CCCAGCAAAG	TTTTTCAGCA	AATGACTAAA	GGTCTCTTTC	10200
TGTTTACCTT	GGTAATGCGC	TTCAAAGACC	TCATGAAAGA	CTTCGTTTTC	GAGAATAAGT	10260
TGCTCGCTTT	GGTTTTGTAA	AATACGGAAA	TTAGGTGCAA	TATCAAGCAG	ATAACCATGT	10320
TTGCCAAGGA	ATTTTGTGT	GAAAGAATCC	ATGGTTCCAA	TGGCAGCGTT	GGGTAGGTCT	10380
GCCAACTGGC	GACCCAAGTG	TTGTTTGAGG	TCGACATCAT	CTGTTTCTTG	GATTTTCTTG	10440
CTGATTTTTT	TCTCTAAACG	TTCTTTAAGT	TCAGTTGCAG	CCTTGACGGT	AAAGGTTGAG	10500
ATAAAGAGTT	GAGAAATTTT	GACACCACGC	GCCAATTGGT	CCAGAAATGCG	CTCTGCCATG	10560
ACAAAGGTCT	TTCCAGAACC	AGCCGATGCT	GAGACCAGGA	TATTCTGGGC	AGAAGTGTAG	10620
ATAGCTTCGA	TTTGCTCGGC	AGTTTCTTTC	TGTTCCCTTG	TCGAATTTGC	TTCTGCTTCT	10680
TGCAGTTTTT	GAATCTCCTC	CTCACTTAAA	AAGGGAATAA	GCTTCATCGA	TTCAACTCCT	10740
CTCTTATTTT	TTCAAGCCAA	GCTTGCTTGA	GTTTTCCTCC	GACCAGACGC	TTGCCATCAG	10800
CTAGGTCCAA	CTTTTCTAGG	AAACGGGCTT	GGCCAGATG	GTAATTGGCT	TCAAAGCCTG	10860
TAATAGCCTG	ATGTTGCTGG	ACGTATGGGG	CAATGCTTCT	GCCATTTTCA	GTATAAGGAT	10920
TGATGGCGAA	CCGGCCTGCT	AAAATCTTCT	CAGCAGCTTT	CTTGTAAGA	TAGGCATTGT	10980
AGTCCAGTAG	GAGCTGAAAT	TCCTCATCTG	TCAGTTGATT	AGCCTTGTTT	TTGTTATAAA	11040

303

ATTGCGCCTAA	ATAACTGCTT	TCTTTTTC	AGAAGAGCCC	TTGGTATTTC	ATAGATTTC	11100
TGGCTTCTAC	CACTGCTCCT	GCCAGACTTT	TTACCGCCAT	CAGAGATTGG	ACAGGTTTCAG	11160
CCATTTCCAA	GTACATGGCG	CCGAAAAAGT	TCTGCTCCCC	TTCTCTTTT	AGGGCAGCAA	11220
GATAGGTTGG	TAACTGAGAA	TTGAGCCCAT	TAAAGAAATG	AGGAAACTGG	AACTGAGTCA	11280
GACTGGATTT	GTAGTCTACT	ACTCCTATCG	CTCCATTAGC	TTTCAAACGG	TCAATCCGGT	11340
CCACCTTGCC	TCGTACAAAG	ACACTGCGTC	CATTGTCTAA	TTGAATAAAG	GCTTGGTCTT	11400
TTCCACCAAA	ATTTGCTTCT	TCTTTGATGG	TTTCGATGGC	TGGATTGTGT	CGGAGAATAT	11460
GTCCAGTTGT	CCGTGCAACA	TCAAGCAAAA	CTTCCTTGGT	AAACTGGGCT	TCCAAACTTT	11520
CTTGATAAAT	AGCTTCAAAT	TCGCGTTCTT	GACTGGTTTC	TTGAATAGCT	TGTTCTAGAC	11580
GTTGGTCAAA	GGAATCTTCA	TTAGGCAACT	GTAAGGCGCG	TTCAAAGATA	CGATGCAAGA	11640
AATTCCTG	ACTACGGGCA	TCAGGATGCA	AACGTAATTC	CTCCTGCAAG	CCTAAACGCT	11700
AGCGTAGGAA	ATAACTGTAT	TCATTGCGAT	AAAACCTCTGT	CAAACCCGAC	GTAGACAGGT	11760
AAAACCTCTG	TTTGGCAGGA	TAGAGAGCTT	GCAAGGTGTC	CTTGGCTAAG	GTCTTGCTGC	11820
TTGGACTGGT	TGGGATAGCT	GGATTTTCCA	GACCTTGCTG	ATCTAGTTTT	TTACCTATGA	11880
CACGCGACAG	AACCTTGACA	AAAGTCAAAT	CTTGCTCAGT	ATCGCTCATC	TCACCCTGCT	11940
GGTGATAGGC	AACCAGACTA	GACAAAAGAC	TGTGATAGGA	CCCCATATCC	TCCTTAGACA	12000
GTCTTTGTG	ATTTCATCTC	TTCTCTCTCC	GCCTAAATCC	AAAATGGATC	AACTCTTGAA	12060
GATAGGCAGA	TTCTTACTT	TCACTTTCGT	TAAAAAGGCT	TGGAGCCGAC	AAGAACAACCT	12120
GCTTACGAGC	AGAATTGACC	AAGGAAAGCA	TAGTGTAGCG	ATTTTCTCTG	AGATTTTCAC	12180
TGCTGGCAAT	CAGTAATTGA	ACGCCTTCTT	CGGTCGCTTG	GTTTAGGTTT	TGCCTTTCTT	12240
CATCTGTCTAG	AAGACTGGTG	TTTTGAGAAA	TTTTTGTTAA	ATTGTCCTGA	GTTAGTCCAA	12300
TAGCATAGAC	AAAGTCAGCA	GTCATGGTG	CAATCAAATC	GTAACCTCTGC	ACCAGAACAG	12360
TGTCCACTGT	TGCTGGAATG	GTACGGTATT	GGGACAAACT	CATTCCAGAA	TGGAGCAAGG	12420
CTAGGAAGTC	TTCCAGACTA	ACCTGTGAAC	CAGCAAAAAC	AGTCGCAAAT	TGTTCTAAAA	12480
CATGGCAGAA	AGCCTTCCAA	ACTTCGGCTT	GTCTTTCCTG	TTCTACAGCT	TCCAAAGTGG	12540
TTGTCAAATC	TTGTAAGTGC	TTGGTCACAG	CTCCTTCTTT	TAGAAAGACA	CTCCATTTTT	12600
GTAGGAGTTT	TTCAGCCTTT	TGTTTTCGGC	TGGCAAAGAG	GGTTTCAAGA	GGTGCTAAAA	12660
TTCTCAGGCG	GAGGACATTC	AAACGCTCAA	GATTAAATTT	TCCATGGTGG	GATTGGGTGA	12720
AGGTTTGCTG	AAAGGCTGGC	AAGCCATTGA	TACCAAGATA	GCGGATATAT	TGCTCAAAAG	12780

304

CATCAATATC AGACTGACTG AGGTCAGTAT ACAAATCAGT TCTAAGAAGA TTAATCAAAT	12840
CCTCCTGACG AAAACGGTAA CGTTTAAAG CTAAATAGA CTCGACAAAC TGAGTCAAGG	12900
GATGATGAGC CATGGCTPCG CTTCTACCAA GATAAAAAGG AATCTGATAC TGGTCAAAAA	12960
TGGTTTTGAG AGATAACTGG TAAGAAGCTA CATCCCCCAA GAGAATACGA AAATGCTTGT	13020
AGCTCAGGTC TGAGTTCTCA TGTAATTTCT GACGAATACT ACGGGCTACT AGCTCCAACT	13080
CCTCCTTTTG CGTCAAACAA GACCAGATTT GTAAATTTTC ACGGTCTTTC TCATCGACAT	13140
CCAAAGCGAG TTCTGAAAAG TCATAAGAAG ACTCCAACAA ACGAGAGGCC TTGTCAAAAC	13200
TATCCATCTT CTCATGACTT TGAGAACAGT CCGGAGCAGG CGTTTGGTAT TTAGAAGCCA	13260
GATGATGGAG AAATTTTACG CTGGCTTGGT AGAGATTGCC CTCGCTAAAA GGAAGTGGTAT	13320
AGGCTTTCTT ACTAGCATAA GCCCCGATAA CAATCTCAAC ACCTTTGCCG TGAAGTAAGT	13380
CCACAACCCG CTCCTCCTCA GCAGAAAAAC GAGTAAAGCC GTCAATGACC AAGCGGATTT	13440
GATTAATAATC ACTACTTACC TTGTCAATTCT CAATAGCCTC AATCAAAATGG GACAACTGAC	13500
TTTCCTGGGC TAACTGACCT TGATTAAGAT AGGCTGTTAC TTTCTCAAAA ATCAAGAGTA	13560
AATCCGCCCT CTTATCCTCA TCTGTTAAAT TCTCCAAGTC CAAAAAATC ATCTGAGATT	13620
TGGTCATCTC ATGGTAAAGC TCAATTAAT GCTGGATCAA TTGAGGATCC TGCTTAATAG	13680
CGCCATAAAC ACGCAAGTCC TTGGGATCGA GTTCGGCAAG GCATTTGTAA AAGGCCAACC	13740
CAAGACCGAT ATCATCAAGA GTAGTTTTAG CTGGTAAATC ATTCAAGACC AGATAGCGAG	13800
CCATTTGAGC AAAGCGCGTG ACGGTAATCG AAAAAGAAGC CTGCTGGGAC AAGTATTCCA	13860
GCACGGCGCG TTCCTTTTCA AAAGAAAGAG AGTTGGGGGC AATGTAGAAG ACCCGCTTGC	13920
CAGCTGCAAC TAGCTCTTCT GCCTCTCTTG TTAGAATTTT TGTCAAAGAA GTCCGAATAT	13980
CAGTATAAAG TAATTTTCATC TCAGCCTCGT TGGAATTTTT CATCACCTA TATTATACCA	14040
TGATTAGCCT CGTAAATCTG TTAAATATT TAGGCCATCC TTTCTTTTCT TCATCATCTG	14100
CTAAATCTTA AATACTTAGC TTTACTTGTA TTAGATAGAA TAAGTCTGGC TACTGAAAAT	14160
CACATAATAA AAAAGCCTCG GTAACAAGGC TTTGAGTTTT ATGATTGTTT CTTAGGTACG	14220
GAATACACTT CAATGTGTG TCCAGTATC TTAATGTCGA CTGGTAGATT GTCTGATTTA	14280
TCGCCATCAA CATCGGACTC TAATTCGATA TCAGAAGAAG TTTTAATATT ACGTGCCTTT	14340
ATATATTCAA TATTCTTGAT AGAATGATTG AACTATAGTA AATTGAAACT ATAATAGTAC	14400
ACCGTGGATG CTAATAATTT TCTAGAAATT AATTTGATTT CCCTAATCAA GCTATTCGTA	14460
TCTTATTTCA ATCTACTATA ATAAAATGAA CCAAAAATAG TACACAATGT GGTATAATCT	14520
TCTTATGGCA TATTCAATAG ATTTTCGTAA AAAAGTCTC TCTTATTGTG AGCGAACAGG	14580

305

TAGTATAACA GAAGCATCAC ACGTTTTCCA AATCTCACGT AATACCATTT ATGGCTGGTT	14640
AAAGCTAAAA GAGAAAACAG GAGAGCTAAA CCACCAAGTA AAAGGAACAA AACCAAGAAA	14700
AGTTGATAGA GATAGACTTA AAAACTATCT TACTGACAAT CCAGATGCTT ATTTGACTGA	14760
AATAGCTTCT GACTTTGGCT GTCATCCAAC TACCATCCAC TATGCGCTCA AAGCTATGGG	14820
CTACACTCGA AAAAAAGAAC CACACCTACT ATGAACAAGA CCCAGAAAAA GTAGCCTTAT	14880
TTCTTAAGAA TTTTAATAGT TTAAAGCACC TAGCACCTGT TTAGATTGAC GAAACAGGAT	14940
TCGATACTTA TTTTATCGA GAATATGGTC GCTCATTAAG AGGTCAGTTA ATAAGAGGCA	15000
AAGTATCTGG AAGAAGATAT CAGAGGATTT CTTTGGTTGC AGGTCTAACA AATGGTGAAT	15060
TAATCGCTCC AATGACTTAC GAAGAGACGA TGACGAGCGA CTTTTTTGAA GCTTGGTTTC	15120
AGAAGTTTCT CTTACCAACA TTAACCACAC CATCGGTTAT TATAGTAAAA TGAAATAAGA	15180
ATAGGGGGGG GGGGGGAGGG GGGGGGAGGG AGA	15213

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6004 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TTATTACCTG AAACATTAAA TTTAATTGGA CATCCCGTTA TCAATTTTAT AATATCATCA	60
AGATTTTTAT TATCTGATTC AGGAATTTTA TCTGATATAA CAACACCATT TTCAAGATAG	120
TTCATTAAAT TATTTGATTC ACTAACATTA GTGTTTGTAT CTCCATCAAG CCAAAAATAA	180
TGGTTATCGG AATCTAAATA CGATGAGTTT AAAATATTAT TACAAATTAT TTGATTTGCT	240
CCACCAGGAA TATATCTCAC TACTAAATTC TGTTTAAGAT TCTCACTACC TGAATGAGTG	300
ATAACAAACT CTAGAATATA TTTAGCTAGT CTATCTTCAA CATAAATCAT CTCCTAGAA	360
TGATACACAT CACCTAATTC AAAAAATGCA TCCTGATAAT CAATATTTTC AATAACATCT	420
ACCTTTTCTC CGTTTTTCAC TAAAAGTTTC ACGGCTTCTC TAGGAAAATC TTTTATAAGT	480
TGTGTAGAAT GTGTAGTGAT AATAATTTGA TGTTTTTTAT TTAACACTC TTGAAGTAAA	540
AACTCTTTAA ATTTATAGAT TGCACTCGGA TGAAGTGAGA TTTCAGGTTC ATCTATTAAT	600
ATTAATGAAT TTGATTGCGC ATTTACTATA TCATTTACTA ACRAAATAAT TCTAGCCTCA	660
CCTGTTCTCG CAAAAGCCTC GGAATATTCT TTTCAGATT TTTTCATCCA AATAGTTTTG	720

306						
GAAGCTTTTA	TATCATCACC	TTTTGAATAC	AAC TTATGTG	TTAAAATTG	AATGCTCTGA	780
TAAGATTCAT	CCATTATTTT	ACTAATAATT	TCACAACTT	TATCATCAAC	TTTAACATTA	840
TCTATAACCA	TTTCCTTTTT	ATAACGCGTA	TAGCTACTTG	TATTATTCCT	TAAAATATCA	900
GCAACTGGCT	TAGATCGTAA	TCTTATAAAA	TCTTGTTTAC	TACGTTGAGT	AGAAATTTTT	960
TTAAAATPAT	AGTGATAGAA	AAATAAATCA	AAAGCAGAAA	CATATTCCTT	ACAATCACAA	1020
AAGACAACAT	TTTTTTCAT	GCCATCCCAT	CTGTCTGTCG	AAGAACTTCC	AATATATTTA	1080
TTTTTGGGTA	ATCTTTCCAT	CTCATATTGT	TTTTGAGGAG	CATATGGTTC	CCAATAATCT	1140
AATCCTTTTT	TTGTTCCAGA	ACGGCCTTTA	AGAACTTCTA	CATTTCTAGA	AGCTTTAATG	1200
TTATAATATG	AATAGATTAA	ACATTGTTTC	CCATCCACTT	CATCTATTTG	ATCAACATTT	1260
GTAATAAAC	AATATTCAGA	CACACTTTTA	TTGGCTGGAG	AACCATATAA	AGCTTGTAAG	1320
ATTGAAGTTT	TATTTACTCC	ATATCTATTA	CAGACACCTC	AGGATTATTT	AACTTATAAG	1380
TTTTAACAGC	TACGGAATCA	ATTTCAACAG	CAACTTGAAC	ATCTATGCCT	GATTTTTTAA	1440
GGCCACTTGT	AGTGCCACCT	GCACCGTTAA	ATAAATCAAT	AGCAACAATT	TTCCCCATAG	1500
TATTCTCCTA	AAGTTTCTCC	TTTTTATTAT	AACATTATCA	AATGTAAAAC	CCAACCCGAT	1560
AGGGTTAGGT	TTTTAACATC	ATTTCAACAA	CTTCTTCATC	TCATCAATAC	GTGCGACGGT	1620
CGCGTCATAT	TTAGCTTGGT	AGTCAGCTTG	TTGTGCGCAT	TCTTTTGGGA	CGACTTCTGG	1680
TTTGCGCTTG	GCTACGAAGC	GTTTCGTTAGA	GAGTTTCTTA	CCAACCATGT	CCAGTTCTTT	1740
TTGCCATTTA	GCAAGTTCCT	TGTCGAGACG	GGCCAGTTCT	TCTTCAACAT	TGAGGAGATC	1800
GGCCAGTGGC	AGGTAGATTT	CTGCTCCTGT	GATGACACTT	GACATAGCCA	GTTCAGGTGC	1860
AGGGATGGTT	GATGCGATTT	CCAAGTGTTC	TGGATTGTGA	AAGCGTTTGA	TATAGTTGAC	1920
ATTGCTGTTA	AAGAAGGCTT	CCAAGTCGCT	ATCGCTTGTC	TTAACAAGGA	TGGTGATAGG	1980
CTTGCTTGGT	GCTACATTTA	CTTCCGCACG	CGCATTCCTA	ACAGCACGAA	TCAAGTCTTT	2040
GAGACTTTCC	ACACCACTGT	GAGCCGCAAG	GTCTTCAAAG	GCTAGATTAA	CAGTTGGGTA	2100
TGCAGCTGTC	ACGATAGAAC	CTTCTGAGAT	TTGTCCAAAG	ATTTCCTCTG	TCACGAATGG	2160
CATGATTGGG	TGAAGGAGAC	GAAGGATCTT	GTCCAGCGTA	TAGAGGAGAA	CAGATCGAGT	2220
AATGACCTTA	TCGTCTTCAT	TGTCGCTGTA	TAGAACTTCC	TTGGTCAACT	CAACATACCA	2280
GTTGGCAAAT	TCTTCCAGTA	TGAAGTTGTA	AAGGATATGA	CCAGCCACAC	CAAACTCGAA	2340
CTTATCAAAG	TTTTCAGTAA	CTTTTGCAAT	GGTTTCGTTG	AGATTGTGGA	GAATCCAGCG	2400
GTCCGTCACA	TTACCAGCCT	CACCTGTTGC	AACTTTTGTG	ACATTGTCAT	GCGCCACATC	2460
CAGCGTCAAA	CCTTCATTGT	TCATGAGGAT	ATAGCGAGAA	ATGTTCCAAA	TTTTGTTAAT	2520

307

AAAGTTCCAT GAAGCATCCA TTTTCTCGTA AGAGAAACGA ACGTCTTGAC CTGGTGCGGA	2580
ACCGTTTGAA AGGAACCAAC GAAGGGCATC AGCACCGTAT TTCTCGATGA CATCCATTGG	2640
GTCAATCCCG TTACCGAGAG ATTTAGACAT CTTGCGTCCT TGCTCGTCAC GGATGAGACC	2700
GTGGATAAGC ACGTTTGGGA ATGGCTGACG ACCAGTAAAT TCCAAGGACT GGAAGATCAT	2760
ACGAGACACC CAGAAGAAGA TGATGTCGTA ACCTGTTACC AAGGTTGAAG TTGGGAAATA	2820
ACGTTTAAAG TCTTCTGAGT CGACTTCAGG CCAGCCCATG GTTGAAAATG GCCAGAGGGC	2880
AGAACTGAAC CAAGTATCCA AGACGTCTTC GTCCTGAGTC CATCCGTCAC CTTCTGGAGC	2940
TTCTTCGCCG ACATACATTT CACCATCAGC ATTGTACCAG GCAGGGATTT GGTGACCCCA	3000
CCAAAGCTGA CGAGAGATAA CCCAGTCGTG GACATTTTCC ATCCATTGAA GGAAGGTATC	3060
GTTGAAACGA GGTGGGTAGA ATTCGACCTT GTCCTCTGTG TCTTGGTTAG CAATGGCGTT	3120
CTTAGCCAAT TGGTCCATCT TGACGAACCA TTGAGTAGAC AAGCGTGGCT CAACTACGAC	3180
ACCTGTACGT TCTGAGTGAC CAACACTGTG GACACGTTTT TCGATTTTGA CAAGGGCACC	3240
GATTTCTTCC AACTTAGCAA CGACTGCCTT ACGAGCTTCA AAACGATCCA TGCCTGAAAA	3300
TTCAAAGGCA AGCTCATTCA TAGTTCGTC GTCGTTTCATG ACGTTGACTT GTGGCAAGTT	3360
ATGACGTTGG CCAACCAAGA AGTCATTTGG ATCGTGGGCA GGTGTGATTT TCACGACACC	3420
AGTACCAAGC TCAGGATCTG CGTGCTCATC TCCAACGATT GGGATGAGTT TATTAGCGAT	3480
TGGAAGGATG ACGTTTMTAC CAATCAAGTC CTGTAGCGC GGGTCTTCTG GATTAACCGC	3540
AACCGCAACG TCCCCAAACA TAGTCTCAGG ACGAGTTGTA GCAACTTCAA GGGCGCGTGA	3600
ACCATCTTCC AGCATGTAAT TCATGTGGTA GAAGGCACCT TCTACATCCT TGTGAATCAC	3660
CTCAATATCA GAAAGGGCTG TGCGAGCTGC TGGGTCCCAG TTGATGATAA ACTCACCACG	3720
ATAGATCCAG CCTTCTTGT AAAGGTTTAC AAAGACCTTA CGAACAGCTT TTGACAAACC	3780
TTATCAAGA GTGAAACGCT CACGAGAATA GTCTACAGAA AGCCCCATCT TGCCCCATTG	3840
TTCTTGATG GTAGTGGCAT ATTCGTCTTT CCATTCCCAG ACCTTCGTCA AGAAAGACTC	3900
ACGACCTAGG TCATAACGCG TAATACCCTC ACCACGTAAG CGCTCCTCAA CCTTAGCCTG	3960
AGTCGCAATA CCAGCGTGGT CCATACCTGG AAGCCAAAGG GTATCAAAGC CTTGCATGCG	4020
TTTTTGACGG ATGATGATAT CCTGCAAAGT CGTATCCCAA GCGTGACCAA GGTGAAGTTT	4080
CCCAGTTACG TTGGTGGTG GAATCACGAT TGAATAAGGC TTAGCCTTTT GATCGCCTGA	4140
AGGCTTGAAA ACATCCGCAT CAAGCCATTT TTGGTAACGA CCAGCCTCAA CCTCGGCTGG	4200
ATTGTATTTA GGTGAAAGTT CTTTAGACAT GTGTGTGTCC TTTCTCTATT TTGTTTATTT	4260

308

TATTTTGAAT	TTGCTTAGCA	GCTTCTTCTG	CAGACAAATT	CGTATTATTT	ATTTTAAAGT	4320
AGTGGTGCAA	CTCATTCGGT	TGATGTTGGG	AATTTAATTG	AAGTGTTTCA	GCGGTCTCTA	4380
AAATTTCTCT	TTCAGATACC	TCAATATGTC	GTTTFAAGGG	TTTGTGCTTT	AATCGATTCT	4440
CCGTTTCGATT	TCGACGTATG	CACTCTTCAA	GACTTGTTTC	CAATTCAACA	AACAGAATCT	4500
CTTGATGAAA	GTTATCCAAT	AAATCCTGAA	TTTGCTTTAA	ATACATCAGC	TGGTACTGAT	4560
TTGAAAAATC	AATTACGTCT	GTTAAAATTA	CTGATCGCTG	ATTTCTTGCA	CTTGCTCCAA	4620
GGAAAGAAAA	GGTAATTCCA	CGAACAAATT	CCCACATCTC	CTCGGTATAA	TCCTGATAGA	4680
TCTCTAGTGC	AAAATCAATG	GCTTGATGGT	TATAAAATAG	GGTAGCATCC	GTCAGTCGAG	4740
ATAATTCTTG	ACCAATGGTC	ATTTTTCCTG	ATGCTGGAGC	ACCAATGATG	AAAAGATGCA	4800
TCAATCACC	TCCCCTCAC	TCCTCAGCAA	GCCATATCTC	AAATCATCAC	AGCAGTTGCC	4860
TTGAGCATCT	TTGCGGTCTC	TTATGCGAGC	TTGAGGGTA	AAGCCAAGCT	TTTCCGAGAC	4920
TCGTTGACTT	TGAAGGTTAT	ATCCAAAGCA	AGTTAGTTCA	ATCTTGTGAA	GACCAAGTTC	4980
TTTAAAGCT	AGATCAATCA	AGGAACACGC	TGCTTCTGGA	ACATAACCTC	GACCCCAATA	5040
GTCTGGGTGC	AAGGTATAGC	CAAGCTCTAG	CACATCATCC	GCATGAAGAT	GGTTGAAGTC	5100
AACAGAACCA	ATGACTTTAT	CGTTTCCTTT	GACGACAATC	CCATAGCCAG	CTGGGAGATT	5160
TTCTTTTGA	GTACGCTCCG	GAAGAATGTG	CTCCAGATAA	TAAATCTCAT	CTTCCAAGAT	5220
CTTGACTGGA	GGAAAACCTG	CTGGATAGGC	GACCTCTGGC	AACTAGCGT	AGGTATGGAT	5280
ATCCTCAGCA	TCCACCACTG	TGCGGACTCG	TAAAACGAGA	CGTTCTGTTT	CGATTTTATC	5340
TGGCAGCTCA	GTTCTTGCCA	TCCTTCTTCC	TCGCTTTTTT	GATGAAACTG	CCCTTCATAT	5400
CTACACGCTT	GTCCAGATAG	CGATAAACGC	GCTGATATCC	ATCTCCCATG	AAATAGGTTG	5460
GGGCAACAG	TTGATTTTGA	AAATGTCCCT	TTTCATCCAG	GAGTTCTGGG	GCAACAAGTC	5520
GCTCAAGAAT	CTTGGCAAAG	ATGTGGCAAA	TACCGTCTTC	CTCAACAATC	CTATCTACCC	5580
GACAATCTAA	AACAAGTGGA	CAGGCGTCTA	AAATAGGAGT	CTGAGTTCGT	TCAGAAATTT	5640
CATAATGCAC	TCCCAAACGT	TCCAATTTCT	CCTGATGACT	GATAAAACCA	GCCTGCTCCA	5700
TCGCAAGCAT	AGAAGTTTCA	TCAGAAATAT	TCACAGTAAA	TTTTTGATAC	TGTTTGATCT	5760
GCTCTGCGGC	ATTCTCTCTC	GCAACGACTC	CAATCACAAC	CCAATCTCCT	AGACTATAAG	5820
AGGAACTACA	GGTCGTGATG	TTATAGCCAA	AATTCTAATC	TTGATATCCT	AAAATAAAAA	5880
CAGGAAAACC	ATAATATAGT	TTACTTGTGT	TAAAAGATTG	CTTCATAACA	ACCCCTTTTG	5940
ACTAAGACGT	AAAAGAAAAG	CCCTGCCATC	TACATGACAG	GGACGAATGT	GTTTATCCGC	6000
GGGG						6004

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5857 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

TGTAGAATTC ACGACAATGC TTCGTTGATT TCTGGGTTGA TTTCGTCGCG TTCTGGCAAG	60
CGAGTCAATG AACCAAAAT AGTACACAAT GTGGTATAAT CCTTTTATGG CATATTCAAT	120
AGATTTTCGT AAAAAAGTTC TCTCTTATTG TGAGCGAACA GGTAAGTATAA CAGAAGCATC	180
ACACGTTTTC CAAATCTCAC GTAATACCAT TTATGGCTGG TTAAAGCTAA AAGAGAAAAC	240
AGGAGAGCTA AACCACCAAG TAAAAGGAAC AAAACCAAGA AAAGTTGATA GAGATAGACT	300
TAAAAACTAT CTTACTGACA ATCCAGATGC TTATTTGACT GAAATAGCTT CTGACTTTGG	360
CTGTCATCCA ACTACCATCC ACTATGCGCT CAAAGCTATG GGCTACACTC GAAAAAGAA	420
CCACACCTAC TATGAACAAG ACCCAGAAAA AGTAGCCTTA TTTCTTAAGA ATTTTAATAG	480
TTTAAAGCAC CTAACACCTG TTTAGATTGA CGAAACAGGA TTCGATACTT ATTTTATCG	540
AGAATATGGT CGCTCATTAA AAGGTCAGTT AATAAGAGGC AAAGTATCTG GAAGAAGATA	600
TCAGAGGATT TCTTTGGTTG CAGGTCTAAC AAATGGTGAG TTAATCGCTC CAATGACTTA	660
CGAAGAGACG ATGACGAGCG ACTTTTGTGA AGCTTGGTTT CAGAAGTTTC TCTTACCAAC	720
ATTAACCACA CCATCGGTTA TTATTATGGA TAATGCAAGA TTCCATAGAA TGGGGAAGCT	780
AGAACTCTTG TGTGAAGAGT TTGGGTATAA ACTTTTACCT CTTCTCCCT ACTCACCTGA	840
GTACAATCCT ATTGAGAAAA CATGGGCTCA TATCAAAAAG CACCTCAAAA AGGTATTACC	900
AAGTTGCAAT ACCTTTTATG AGGCTTTTTT GTCTTGTTCT TGTTFCAATT GACTATATAA	960
ATTGTCTAAG CGAAACAACC GATAAGAATT GGCACAAAAG CGACCGTATT TTTGTTACCA	1020
ATACAGGAAA AACAGTTCAT AGTTCTATCT TGAGCAAGTC TCTCCAGCGA GCAAACGAAC	1080
GCCTTAAAAA ACCAATTCCC AAACATCTGT CCCCTCACAT CTTCAGACAC ACCACTATTA	1140
GCATCTTATC AGAAAATAAA ATTCCTTTAA AAACAATCAC GGACAGGGTT GGTCTATCCG	1200
ACTCTGAAGT CACTACTTCC ATCTACACCC ACGTCACAAA GAACATGAAA GATGAAGCAA	1260
TCAATGTAAT GGATAAAGTT ATGAAAAAGA TTTTAAAAA AGTTTGTGCC CTTTTTGCC	1320
CTCTAAATAC AAAAATAGCC CTTCCGATAA AATCCGAGGG GCTAGAAACG TTGTTAAATC	1380

310

AACGGCCGAA CTTTGAATT TCATGGTTCG GGATAAAATA GTTCACTGAA CTATTTTATT	1440
TTTAAAGGTT ATCATAATAT CAAATAGTTC AATTAAATAC GCTAAATTAC TAATATACTT	1500
TTTACCTTTT TCATTCTAAA ATGTAAAGTA CAAACAATTA CAATATACTA GAGGGGGAGT	1560
AAAAAAGGTA TTAAATCGAT GAGTTCAGCA GGCAAGAAAA TAGCACCTTT ACGGGTGCTA	1620
TTTTTTAATT AACGCCACGT TAACTTTTGA TTGATGAATT TTATTGTTTG GCACTTCTTT	1680
CATTTACCGG TAAACATCGA TGAAATTCCT TCCAACATTA TTTTGGAGT TAACTGCATT	1740
TATTTTGTGA TTAATAACTT TTTTAGTATC GAAAGAATGG TTTAAGAAAT CCATAACTAA	1800
CTCTCCTTTC TCATCCTGTA ATCAAGATTT TTATCAATGT CAAAATAGTA TTTTCTATCA	1860
ATCCAAATTG GTCCTTCTCC TTTAGAAATA GCAAGTACAT CTACCGGACC TCCTACTGTT	1920
TCAAGAGTGT TGACAATTTT TCTCTTAAAT GAAGTTAATT CAATAAATGT TTTAGCTGTA	1980
CTCGCCATTT CATTAAGTGG TTGCATTCCA ATAAGGTCTA TTATAGGATT TATATAATAT	2040
TTTGTCTGTA TAGATGATAT ATTTTCAAAT ATATTCTCAA TTTCATCACC CAATCCATTT	2100
TTCTCCATAA CTGATGATAC TTGCTCTGCG ATATATACAT TTAAGTTAGG ATCTATACCA	2160
TTCATAATCG TCTCAACCAT CTCTGACTGT GCAAAAGGGA TTATATGACA AGTTTATGA	2220
TGATTTATCA CACTTTCATT AATAACTTTC CAAATTAATC GTTTAGAAAA AATCCATAT	2280
AATTCAATTT GTCTTATAGA TGGAAATATC TCGTCTGTAC CATAACCTGC TATAACTAAT	2340
CCAGTTATGT TTGTTGAGTC ATATCCAATG AAAATCGCTT TATATAAAGA TTTAGCAATA	2400
ACTTCAACCT CATCATCAGT ATGAGGAAAG GATTTAAAAA CATCGTCTAC AATGCTTTTT	2460
ATTAACCTTA ACTCAGCTTC AAAAAATTC AATTACTTT CAGCTTCTAC TTTGAAATT	2520
TCTAAACTAA AATTAGTTAT AGCATTTAAT AAAATTTTAT TAAAATCATC TAGAGTGATG	2580
GTTTCACCAT TAGAACTCT TAAATCAGCT GTTCTTGGC CTTCATAGGC AATGCTGTCC	2640
AAAATACTTC TTGTACTTCT GACAATATAA TTTCTTAATA AATCCTCAAC TTGTAGATGT	2700
TTAAAGGAAA TTAAAAATTC TATTAGCTTT TCAACGTATT GGGCAGTATT ATCTAATAAA	2760
TCTGTGCCAA TAGCCTGCTT AACTCATTT AAAATTACCT CCCACGGAAT TTCCATAAAC	2820
GAAGCGTTCC CATATATCAT GATCCCCACG GAATGTTCTT TTGATAAAGT GAATAATTTT	2880
CGGGCGCTAT TAAAACTTT TGAAATTTTC CCGTCTGATA AGGTTACAGC GCTATCAGAA	2940
GCCAATACAA CACCATTTTT ATTTAATATT CCAATTTCTG CTGTCAAAAT ATCACCTAAA	3000
CTTCTAAAC CTGCTCATGC TCTAATGGTA CAACAGCTAA GGTCTTACCA AGACTTGCCA	3060
ACACTTTTAA TACTGTATCA AGTTGTGGC TTGTCTTTCC TGTTTCCATT CTAGCGATAA	3120
CTGGCTGACT AACACCGCTC ATCTCCTCTA GTTCTTCTG ACTAATACCC TTTTCATTTT	3180

311

TAGCCTCGAT AAGCTCACTC ATGATAGCCA CGCGCATATC ACTTTCCAAA ATTTCTCTTT	3240
TGCTGAATAA TTCAGCTCTT ACATCTTTCC AGTTACTACC AATAGCATTA TTTTTCATTG	3300
TCTAAACCTC TTTCTTTTAA ATCTGCAAGT TCACGTTTAG CTTGCTCAAT CTCTCTTTTG	3360
GGTGTCTTCT GTGTCCTTTT CATAAAATGA TGCAGTAAAA CAAACTACC ATCCATCCAA	3420
GCAACAAATA AAATTCATC TCTAAGTGGT CTCAGCTCCC AAATTCAGC ATCTAAATGC	3480
TTAATATATG GTTCGCCTGC GCGTGTCCA TGTGGCTTA ACAACTCAAT ATAATCATTA	3540
ATTTTATTA GCTTAATCTT GCTATCTTTC CCTTTTTTAC TGGTAAGCTC TCGCATATAA	3600
TCAAAAACAG GCTCATTGCC GTTTTATCC TTGTAAAAAT AGATATTATG CACTATTAAC	3660
ACCTCTTCCT AATAACAATT ATAACCTAAA AGTTATTGTT TGTAATACT TTTAAGTTAT	3720
TAAATAAAAA AGCACCTAGT TTCCTAGATG CTAGCACAAT GACACGGATT CGCACCGTGG	3780
CTACCTCTAT CAAGGTGTAC TCCTCTATA CTATCCCTTG TGCTTTAGAA TATTATACCA	3840
CACAATCAAC TAGATACCTA CCATCTCATG ATATACCCCC ATTTTGGGCA AGGGTACAAC	3900
GCTAAAATAC AANTCAGAAT AGATATTAAA CCACTTATTT AACTTATCAT AAGCTGGTGA	3960
TTGACTGATA AATAATATCC GCTGACAAGC TCCGATAACA TTCATGTGAT TGTACACATA	4020
AACCTCTTTT ACAGCCTCTA AAATGTCAGC CTCACCTGTT TGTACCCTAA TATCTGTTAT	4080
CTGCTTGATA GTTGCGTATT TTTGATAAGC TAGCATATCT TGATTTTTAG CAGCATCAAA	4140
CATTTTACGC TCAAGGACAC TATACCTAGG TTGTTCTTTA TCTCGCATGA AATACCACTT	4200
GAGCCATAAA ATCTTTCTC GGTGTATTAC AGAAATACGC TCAATTTTCT TCTTTGTCAT	4260
TGCTACCTCC TAAATCATCA ATTTAACAAT TCTAACCCT CACTTTTAGA AATAGTTGCA	4320
TAGATCTTGT TCGATGTATG ATACAAAGGT TCTAAATCTT TTTCCACCCT AATATAGTTC	4380
ATCTTATCCT CATGAGTAGG AAAGTATAGT ATTTCCGTTT CATCCTCGTT TAGGATACGA	4440
TTGCACCAAT CATCAATAAT AACTGGCACT TCCCCTCAC GCCATTTTTT AAGGTTTTCT	4500
AAAAGTTCAT TATCACTAAA TAGCTCGCCA TCTATTGGA AAAATTCCCC TAAGTCATTG	4560
TTTCCTTCAA CAATAATAAA CTCTGGCATA TTTCTATTAC TTAATAACTC CTTGAGTTCT	4620
TGTAACCTTT TGATTTCTTT TAGATACTTC CTCAATTTCC AACCTCAATT CTTCAATCTG	4680
CCTTACTACT CCAAAAATTT CATGGTCTTT ATAAGATTGT TCAAGTATAG CCTTTGCTGC	4740
TTGAGTTCTT ATAAACGGGT TGACCTTACT GTCCATCATA ATATCATTGA GTACAGAAAC	4800
AGCGTTAGAT GATGCTAAAT AAAGCATTG AGTTGTTTTA TCCATCATCT CATCTTGCTT	4860
TATCCTCAAT GTCTTTTAA CCGCTGCAAC TTTTAGATAC TTATGACCTG TTGCGCGTGA	4920

312

TACCCCTGCT TTTTGACATG CTTTGTCTAT CGTTGGCTCG GTAAGCATGG CATCTATGAA	4980
TTTAATTTGC TTGGACGTAA GGTATCATT TTCATTTCTT GCCATCTATT ACCTCCTCAT	5040
TATCAAAATA AAGGGTTGCC CCTTTATTTT CCTATGCTAG ATAATTCTGC AATTCTGCAT	5100
CCATTGCCTC TGAATTGCCC TCAACAATCA TTTCATGCTG TACTAAATCA ATCTTATCTC	5160
CGTTAATAAG TAAACCACCG TGGAAATAAT CAATTTTCTT ATCAAGGAAA TGTACTAGCT	5220
TTTCAAGGCG TTGCTGTTGG CTGAATTGCT CCATGTCAAT TTCGATATAA GCAAGGGTAG	5280
TATCATATATC CATAATATCT TCTAATTTT TAAGAGCTAG AGGTTTATTT TTATATTTTT	5340
CTAGGTATTC TCTCATTTCT GCCACTGTTA ATTTGATACT AGATAATAAA CTAGTTCAG	5400
CTGCATCATC TGCTGTAATA GGCTCTTCTT TTGATTCATG GTTTGCTAGT TCAGCATTTT	5460
TCTCTTTTTC TAGTTGCTGA TACAATAGCT GAGCAGTATT TTGGGAATAG TTTTCGCCCT	5520
CTTTTTTATA TTTTAAAAGT TCTTGCTCTG CATACACTTT CCCGATAATC ACTTCCTTAT	5580
AAACTAATTG CCCATCTTGA GCTTTTAGCT TAATACTCCC ATGCTCTGGA ATTTCAATAT	5640
ACTTAATTAT ACCATTTTTT GAGTATAAAA CAAAGCCTTT CTCCATCATT TTTAATAATT	5700
TATCATCCTT GTTTTCAGTC ATGCTTTTCT CCTTTATTTT ATTTTATTAT AATCTGAATA	5760
CCCCTAGTCT ATTTATTTCA CTAGGTTTTT AGGGTTCGTA TGCTAAAATA CTACCCTTTT	5820
TGTGTACCTT ATGGCTGACT TTTCAAATTG GTTAGTT	5857

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10254 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AAAATGATAG CAGGAGAGTT TTCCCGTCCA TCAGACCCAG AACTGAGAGC CTTAGCTCAG	60
GCTTCTCGCC AAAACAGGC CGCCTTTAAC AAGGAAGAGA ACCCCTTGAA GGGAGCCGAA	120
ATCATCAAGA CTGGTTTGC CTCAACCGGG AAAAATCTTT ACATCAACAC TCGCTTGATG	180
GTGGACTACG GTGTCAACAT CCATCTAGGG GAAAATTTTT ATTCTAATTG GAACTTGACC	240
ATGCTGGATA TCTGTCCCAT TCGTATCGGG GACAATGCTA TGATTGGTCC TAATTGTCAG	300
TTTTTGACAC CCCTCCATCC ACTAGATCCA CAGGAACGCA ATTCAGGTAT CGAGTACGGA	360
AAGCCTATCA CAATCGGAGA TAATTTCTGG ACTGGTGGTG GCGTCATTGT CCTTCCTGGA	420
GTGACACTGG GAAATAATGT CGTGCAGGA GCAGGGGCAG TAATTACCAA ATCTTTTGGC	480

GACAACGTTG TCCTAGCTGG CAATCCTGCG CGCGTGATTA AGGAAATACC TGTAAATAG	540
AAGTAAAAAG GAACAGCTGG GGTGTTTCT TTTTGTAGG TTTCATCATT TTTTACCCAG	600
TTACATTTA CCTACTCTAT CTCTTAGCAA GTCTGTTTCA TTAAGCAAGT TCAAAGCATC	660
TCGTAAGTGG GATGTTTTTC TCCTCAGTTC ATCAGCTTCC TCCTTGACAC TCGGTCAGAT	720
TTTGATACAA TAGTACAAAA TTAGAGGAGG CAGGCTATGA TTCAGAAACA TGCGATTCCT	780
ATTTTAGAGT TTGATGACAA TCCTCAGGCG GTTATCATGC CCAATCACGA GGGGCTGGAC	840
TTGCAGTTGC CAAAGAAGTG TGTATTATGCA TTTTATAGTG AGGAGATTGA CCGCTATGCG	900
AGGGAAGTAG GGGCGAAGTG TGTGGCGAA TTTGTTTCTG CCACCAAGAC CTATCCAGTT	960
TATGTCGTGA ACTACAAGGA CGAGGAGGTC TGTCTGGCTC AGGCTCCTGT TGGCTCCGCT	1020
CCAGCAGCCC AGTTTATGGA TTGGTTGATT GGCTATGGTG TGGAGCAGAT TATCTCTACT	1080
GGGACCTGTG GTGTCTAGC TGATATAGAG GAAATGCCT TTCTAGTCCC GTTTCGCGCT	1140
CTGCGAGATG AAGGAGCCAG TTACCACTAT GTGGCACCTT GTCGTTATAT GGAAATGCAG	1200
CCAGAGGCTA TTGCTGCTAT TGAGGAAGTT TTGGAAGACA GAGGGATTCC TTATGAAGAA	1260
GTCATGACCT GGACGACAGA CGGTTTTTAC CGAGAAACGG CTGAAAAGGT GGCTTATCGT	1320
AAGGAAGAAG GCTGTGCTGT TGTGGAGATG GAGTGTCTG CTCTTGCGGC AGTAGCTCAA	1380
TTGCCGTGGG TTCTCTGGGG TGAATTGTTG TTCACAGCAG ATTCTCTAGC GGACTTGGAC	1440
CAGTACCACA GTCGTGACTG GGGCTCGGAA GCTTTTAATA AGGCGCTAGA ACTGAGTTTA	1500
GCAAGTGTTC ACCACCTTTA GTTGTAAGTG CAAAGGATTT GTTTTATCAT AAAATGTCTA	1560
GCTCATACTT TTCAAAAATA TGTTTAAACG AGGTCACCTT CCTCTGTGCC TAGGCATGTT	1620
GAGGTGGGA AAAATCTTTA AAATCAGAAA AACGTATCAT ATCAGGTGAT GAAAACCTTG	1680
ACACTATGCG TTTTATGTCG ATAAGATTTA GAGTGAGATG AAATGATACT CTTCGAAAA	1740
CTCTTCAAA CAGGTCAGCT TCACCTTGCC GTAGGTATAT GTTACTGACT TCGTCAGTCT	1800
TATCCGGCAA CCTCAAAACG GTGTTTGTAG CTGACTTTCGT CAGTTCTATT TGCAACCTCA	1860
AAACAGTGT TTGAGCAACC TGTGACTAGC TTTCTAATCG ATGCCCTGGT TTTTATTGCC	1920
TATAATCAAA AAGAGAAATT TTCTCCTGAA AAGCATATAG AGTAGCTGGC GTTAAAAGCT	1980
CCTGTCTTGC TTTTGTGACC TATAGTCACA TCTATCAAGT ATTGTTCTTG CCTAAGCTAT	2040
CAATAAAAAG GTGGCATTTT TTAGGCTTGG TGTAGTAGA TTTTGCCCTA TCCTATCTAA	2100
GTCATTTTGA ACTTTTATG GTACAATGGA AACATGTTAT TCAAATTATC TAAGGAAAAA	2160
ATAGAGCTAG GCTTATCTCG TTTATCGCCA GCCCGTCGTA TTTTGTGAG TTTTGCCTTG	2220

314

GTCATTTTAC TAGGCTCTCT TCTTTTGAGC TTGCCCTTTG TCCAAGTTGA AAGCTCACGA	2280
GCGACTTATT TTGATCATCT TTTCACCTGCT GTCTCTGCAG TCTGTGTGAC GGGTCTCTCA	2340
ACCCTTCCAG TAGCTCACAC CTATAATATC TGGGGTCAAA TAATCTGTTT GCTCTTGATT	2400
CAGATCGGTG GTCTAGGGCT CATGACCTTT ATTGGGGTTT TCTATATCCA GAGCAAGCAA	2460
AAGCTTAGTC TTCGTAGCCG TGCAACTATT CAGGATAGTT TTAGTTATGG AGAAACTCGA	2520
TCTTTGAGAA AGTTTGTCTA TTCTATTTT CTCACGACCT TTTTGGTTGA GAGCTTGGGA	2580
GCTATTTTGC TTAGTTTTCG CCTATTTCCT CAACTTGGCT GGGGACGTGG TCTTTTTAGT	2640
TCCATTTTTC TAGCGATCTC AGCCTTCTGT AATGCCGTTT TTGATAATTT AGGGAGCACC	2700
AGTTTATTTG CTTTTCAGAC CGATTTACTG GTCAATCTGG TGATTGCAGG CTTGATTATT	2760
ACAGGCGGCC TTGGTTTAT GGTCTGGTTT GATTTGGCTG GTCATGTAGG AAGAAAGAAA	2820
AAAGGACGTC TGCACTTTCA TACGAAGCTT GTACTATTAT TGACTATAGG TTTGTTGTTA	2880
TTTGGAACAG CAACTACTCT CTTTCTTGAG TGGAAACAATG CTGGAACGAT TGGCAATCTC	2940
CCTGTTGCCG ATAAGGTTT AGTTAGCTTT TTTCAAACAG TGACGATGCG AACAGCTGGC	3000
TTTTCTACGA TAGATTATAC TCAGGCTCAT CCTGTGACTC TTTTGATTTA TATCTTACAG	3060
ATGTTTCTAG GTGGGGCACC TGGAGGAACA GCTGGGGGAC TCAAGATTAC GACATTTTMT	3120
GTCTCTTGG TCTTTCACG AAGTGAGCTT CTAGGCTTGC CTCATGCCAA TGTTCGAGAG	3180
CGAACGATCG CGCCGCGAAC GGTTCAAAA TCCTTTAGTG TCTTTATTAT CTTTTTGATG	3240
AGCTTCTTGA TAGGATTGAT TCTGCTAGG ATAACAGCCA AAGGCAATCC TCCCTTTATC	3300
CACCTCGTAT TTGAAACCAT TTCAGCTCTT AGTACAGTTG GTGTAACGCG AAATCTGACT	3360
CCTGACCTTG GGAATTTGGC TCTCAGTGT ATCATGCCAC TTATGTTTAT GGGACGAATT	3420
GGTCCCTTGA CCTTGTGTGT TAGCTTGGCA GATTACCATC CAGAAAAGAA AGATATGATT	3480
CACTATATGA AAGCAGATAT TAGTATTGGT TAAGAAAGGA AAGAGCATGT CAGATCGTAC	3540
GATTGGAATT TTGGGCTTGG GAATTTTGG GAGCAGTGT CTAGCTGCCC TAGCCAAGCA	3600
GGATATGAAT ATTATCGCTA TTGATGACCA CGCAGAGCGC ATCAATCAGT TTGAGCCAGT	3660
TTTGGCGCGT GGAGTGATTG GTGACATCAC AGATGAAGAA TTATTGAGAT CAGCAGGGAT	3720
TGATACCTGC GATACCGTTG TAGTCGCGAC AGGTGAAAAT CTGGAGTCGA GTGTGCTTGC	3780
GGTTATGCAC TGTAAGAGTT TGGGGGTACC GACTGTTATT GCTAAGGTCA AAAGTCAGAC	3840
CGCTAAGAAA GTGCTAGAAA AGATTGGAGC TGACTCGGTT ATCTCGCCAG AGTATGAAAT	3900
GGGGCAGTCT CTAGCACAGA CCATTCTTTT CCATAATAGT GTTGATGTCT TTCAGTTGGA	3960
TAAAAATGTG TCTATCGTGG AGATGAAAAT TCCTCAGTCT TGGGCAGGTC AAAGTCTGAG	4020

TAAATTAGAC CTCCGTGGCA AATACAATCT GAATATTTTG GGTTCCTGAG AGCAGGAAAA	4080
TTCCCCATTG GATGTTGAAT TTGGACCAGA TGACCTCTTG AAGCAGATA CCTATATTTT	4140
GGCAGTCATC AACAACCAGT ATTTGGATAC CCTAGTAGCA TTGAATTCGT AAAGAGGGAT	4200
GACCCCTCTT TTTTGATGCC TAAGATGGCA AATAGAGACA GAAGCCCTT GTCTTCTAGT	4260
AAAAGTTCTT CAAAGGCTGG ACTTTATGGT AAAATAGAAA GAAGTGACAA GAGAGAGTAA	4320
TACTCAATGA AAATCAAAGA TCAAACTAGG AACTAGCTA CGGGCTGCTC AAAACACTGT	4380
TTTGAGGTTG CAGATAGAAC TGACGAAGTC AGTAACATCT ATACGGCAAG GCGACGTGA	4440
CGCGGTTTGA AGAGATTTTC GAAGAGTATA AGAAAAATC AGTCCCTTAA AGGAGTAGAT	4500
TATGAAGTTA TTGTCTATCG CAATTCTAG CTATAATGCA GCAGCCTATC TTCATTACTG	4560
TGTGGAGTCG CTAGTGATTG GTGGTGAGCA AGTTGGGATT TTGATTATCA ATGACGGGTC	4620
TCAGGATCAG ACTCAGGAAA TCGCTGAGTG TTAGCTAGC AAGTATCCTA ATATCGTTAG	4680
AGCCATCTAT CAGGAAAATA AATGCCATGG CGGTGCGTC AATCGTGGCT TGGTAGAGGC	4740
TTCTGGGCGC TATTTTAAAG TAGTTGACAG TGATGACTGG GTGGATCCTC GTGCCTACTT	4800
GAAAAATCTT GAAACCTTGC AGGAAGTTGA GAGCAAAGGT CAAGAGGTGG ATGCTCTTGT	4860
GACCAATTTT GTCTATGAAA AGGAAGGGCA GTCTCGTAAG AAGAGTATGA GTTACGATTC	4920
ACTCTTGCCCT GTTCGGCAGA TTTTGGCTG GGACCAGGTC GGAATTTCT CCAAAGGCCA	4980
GTATACCATG ATGCACTCGC TGATTATCG GACAGATTG TTGCGTGCTA GCCAGTTCTA	5040
ACTGCCTGAA CATACTTTTT ATGTCGATAA TCTCTTGTC TTTACGCCCC TTCAGCAGGT	5100
CAAGACCATG TACTATCTGC CTGTCGATT CTATCGTTAT TTGATTGGGC GTGAGGACCA	5160
GTCTGTCAAT GAGCAAGTGA TGATTAAGTG CATTGACCAG CAACTCAAGG TCAATCGACT	5220
CTTGATAGAC CAACTTGATT TGTCCCAAGT GAGTCATCCC AAAATGCGAG AATATCTGCT	5280
GAATCATATT GAACTCACGA CGGTGATTTC CAGTACCCTG CTCAACCGAT CTGGAACAGC	5340
GGAGCATCTG GCAAAAAAAC GCCAATTGTG GACCTATATT CAGCAGAAAA ATCCAGAAGT	5400
CTTTCAGGCT ATTCGTAAGA CCATGTTGAG CCGTTTGACC AAACATTCTG TCTTGCCAGA	5460
TCGCAAACTG TCCAATGTCG TCTATCAAAT CACCAAATCT GTTTATGGAT TTAATTAATA	5520
TAAGTGTTTT ATAAGAGGGA TTTAAGAAAA ATTTTAACTT TTTCTTAGTC CTTTTTAATT	5580
TCAGGAGATT ATACTAGAGT CATCAAATAA AGAAAGACTC TAAGGAGAAT CCTATGAAAT	5640
TCAATCCAAA TCAAAGATAT ACTCGTTGGT CTATTCGCCG TCTCAGTGTC GGTGTTGCCT	5700
CAGTTGTTGT GGCTAGTGGC TTCTTTGTCC TAGTTGCTCA GCCAAGTTCT GTACGTGCCG	5760

316

ATGGGCTCAA	TCCAACCCCA	GGTCAAGTCT	TACCTGAAGA	GACATCGGGA	ACGAAAGAGG	5820
GTGACTTATC	AGAAAAACCA	GGAGACACCG	TTCTCACTCA	AGCGAAACCT	GAGGGCGTTA	5880
CTGGAAATAC	GAATTCACTT	CCGACACCTA	CAGAAAGAAC	TGAAGTGAGC	GAGGAAACAA	5940
GCCCTTCTAG	TCTGGATACA	CTTTTGTAAA	AAGATGAAGA	AGCTCAAAAA	AATCCAGAGC	6000
TAACAGATGT	CTTAAAAGAA	ACTGTAGATA	CAGCTGATGT	GGATGGGACA	CAAGCAAGTC	6060
CAGCAGAAAC	TACTCCTGAA	CAAGTAAAAG	GTGGAGTGAA	AGAAAATACA	AAAGACAGCA	6120
TCGATGTTCC	TGCTGCTTAT	CTTGAAAAAG	CTGAAGGGAA	AGGTCTTTC	ACTGCCCGTG	6180
TAAACCAAGT	AATTCCTTAT	GAATATTTCG	CTGGTGATGG	TATGTTAACT	CGTCTATTAC	6240
TAAAAGCTTC	GGATAATGCT	CCTTGGTCTG	ACAATGGTAC	TGCTAAAAAT	CCTGCTTTAC	6300
CTCCTCTTGA	AGGATTAACA	AAAGGGAAAT	ACTTCTATGA	AGTAGACTTA	AATGGCAATA	6360
CTGTTGGTAA	ACAAGGTCAA	GCTTTAATTG	ATCAACTTCG	CGCTAATGGT	ACTCAAACCTT	6420
ATAAAGCTAC	TGTTAAAGTT	TACGGAAATA	AAGACGGTAA	AGCTGACTTG	ACTAATCTAG	6480
TTGCTACTAA	AAATGTAGAC	ATCAACATCA	ATGGATTAGT	TGCTAAAGAA	ACAGTTCAAA	6540
AAGCCGTTGC	AGACAACGTT	AAAGACAGTA	TCGATGTTCC	AGCAGCCTAC	CTAGAAAAAG	6600
CCAAGGGTGA	AGGTCCATTC	ACAGCAGGTG	TCAACCATGT	GATTCCATAC	GAACCTCTCG	6660
CAGGTGATGG	CATGTTGACT	CGTCTCTTGC	TCAAGGCATC	TGACAAGGCA	CCATGGTCAG	6720
ATAACGGCGA	CGCTAAAAAC	CCAGCCCTAT	CTCCACTAGG	CGAAAACGTG	AAGACCAAAG	6780
GTCAATACTT	CTATCAAGTA	GCCTTGGACG	GAAATGTAGC	TGGCAAAGAA	AAACAAGCGC	6840
TCATTGACCA	GTTCAGGACA	AAyGGTACTC	AAACTTACAG	CGCTACAGTC	AATGTCTATG	6900
GTAACAAAGA	CGGTAAACCA	GACTTGGACA	ACATCGTAGC	AACTAAAAAA	GTCACTATTA	6960
ACATAAACGG	TTTAATTTCT	AAAGAAACAG	TTCAAAAAGC	CGTTGCAGAC	AACGTTAAAG	7020
ACAGTATCGA	TGTTCCAGCA	GCCTACCTAG	AAAAAGCCAA	GGGTGAAGGT	CCATTACACG	7080
CAGGTGTCAA	CCATGTGATT	CCATACGAAC	TCTTCGCAGG	TGATGGTATG	TTGACTCGTC	7140
TCTTGCTCAA	GGCATCTGAC	AAGGCACCAT	GGTCAGATAA	CGGTGACGCT	AAAACCCAG	7200
CCCTATCTCC	ACTAGGTGAA	AACGTGAAGA	CCAAAGGTCA	ATACTTCTAT	CAATTAGCCT	7260
TGGACGGAAA	TGTAGCTGGC	AAAGAAAAAC	AAGCGCTCAT	TGACCAGTTC	CGAGCAAACG	7320
GTACTIONAAC	TTACAGCGCT	ACAGTCAATG	TCTATGGTAA	CAAAGACGGT	AAACCAGACT	7380
TGGACAACAT	CGTAGCAACT	AAAAAAGTCA	CTATTAAACAT	AAACGGTTTA	ATTTCTAAAG	7440
AAACAGTTCA	AAAAGCCGTT	GCAGACAACG	TTAAGGACAG	TATCGATGTT	CCAGCAGCCT	7500
ACCTAGAAAA	GGCCAAGGGT	GAAGGTCCAT	TCACAGCAGG	TGTCAACCAT	GTGATTCCAT	7560

317

ACGAACTCTT CGCAGGTGAT GGCATGTTGA CTCGTCTCTT GCTCAAGGCA TCTGACAAGG	7620
CACCATGGTC AGATAACGGC GACGCTAAAA ACCCAGCTCT ATCTCCACTA GGTGAAAACG	7680
TGAAGACCAA AGGTCAATAC TTCTATCAAG TAGCCTTGGA CGGAAATGTA GCTGGCAAAG	7740
AAAAACAAGC GCTCATTGAC CAGTTCCGAG CAAACGGTAC TCAAACCTAC AGCGCTACAG	7800
TCAATGTCTA TGGTAACAAA GACGGTAAAC CAGACTTGGA CAACATCGTA GCAACTAAAA	7860
AAGTCACTAT TAAGATAAAT GTTAAAGAAA CATCAGACAC AGCAAATGGT TCATTATCAC	7920
CTTCTAACTC TGGTTCTGGC GTGACTCCGA TGAATCACAA TCATGCTACA GGTACTACAG	7980
ATAGCATGCC TGCTGACACC ATGACAAGTT CTACCAACAC GATGGCAGGT GAAAACATGG	8040
CTGCTTCTGC TAACAAGATG TCTGATACGA TGATGTCAGA GGATAAAGCT ATGCTACCAA	8100
ATACTGGTGA GACTCAAACA TCAATGGCAA GTATTGGTTT CCTTGGGCTT GCGCTTGACG	8160
GTTTACTCGG TGGTCTAGGT TTGAAAAACA AAAAGAAGA AAATAATCA GCTAAGGAAA	8220
TAAATGATGG ATAGTGGGCT GACTAAGATT AGTTTAAACA CTCAATCAGC AATCAGGACT	8280
TTCTTTCAAT AGCAGATTAA AATCATCGTA AAACAATAAA AATAGTGTTA TACTTAAAGC	8340
AGTATAGCAC TGTTTTTATC AAAGGAGAGA CAGATGGGAA AGACAATTTT ACTCGTTGAC	8400
GACGAGGTAG AAATCACAGA TATTCATCAG AGATACTTAA TTCAGGCAGG TTATCAGGTC	8460
TTGGTAGCCC ATGATGGACT GGAAGCGCTA GAGCTGTTCA AGAAAAAACC GATTGATTTG	8520
ATTATCACAG ATGTCATGAT GCCTCGGATG GATGGTTATG ATTTAATCAG TGAGGTTCAA	8580
TACTTATCAC CAGAGCAGCC TTTCTATTT ATTACTGCTA AGACCAGTGA ACAGGACAAG	8640
ATTTACGGCC TGAGCTTGGG AGCAGATGAT TTTATTGCTA AGCCTTTTAG CCCACGTGAG	8700
CTGGTTTTGC GTGTCCACAA TATTTGCGC CGCCTTCATC GTGGGGGCGA AACAGAGCTG	8760
ATTTCCCTTG GCAATCTAAA AATGAATCAT AGTAGTCATG AAGTTCAAAT AGGAGAAGAA	8820
ATGCTGGATT TAACTGTTAA ATCATTGAA TTGCTGTGGA TTTTAGCTAG TAATCCAGAG	8880
CGAGTTTTCT CCAAGACAGA CCTCTATGAA AAGATCTGGA AAGAAGACTA CGTGGATGAC	8940
ACCAATACCT TGAATGTGCA TATCCATGCT CTTCGACAGG AGCTGGCAAA ATATAGTAGT	9000
GACCAAATC CCACTATTAA GACAGTTTGG GGGTTGGGAT ATAAGATAGA GAAACCGAGA	9060
GGACAAACAT GAAACTAAAA AGTTATATTT TGGTTGGATA TATTATTTCA ACCCTCTTAA	9120
CCATTTTGGT TGTTTTTTGG GCTGTTCAAA AAATGCTGAT TGCGAAAGGC GAGATTTACT	9180
TTTTGCTTGG GATGACCATC GTTGCCAGCC TTGTCGGTGC TGGGATTAGT CTCTTTCTCC	9240
TATTGCCAGT CTTTACGTCG TTGGGCAAAC TCAAGGAGCA TGCCAAGCGG GTAGCGGCCA	9300

318

AGGATTTTCC TTCAAATTTG GAGGTTCAAG GTCCTGTAGA ATTTTCAGCAA TTAGGGCAAA	9360
CTTTTAATGA GATGTCCCAT GATTTCAGG TAAGCTTTGA TTCCTTGAA GAAAGCGAAC	9420
GAGAAAAGGG CTTGATGATT GCCCAGTTGT CGCATGATAT TAAGACTCCT ATCACTTCGA	9480
TCCAAGCGAC GGTAGAAGGG ATTTTGGATG GGATTATCAA GGAGTCGGAG CAAGCTCATT	9540
ATCTAGCAAC CATTCGACGC CAGACGGAGA GGCTCAATAA ACTGGTTGAG GAGTTGAATT	9600
TTTTGACCCT AAACACAGCT AGAAATCAGG TGGAAACTAC CAGTAAAGAC AGTATTTTTC	9660
TGGACAAGCT CTTAATFGAG TGCATGAGTG AATTTTCAGTT TTTGATTGAG CAGGAGAGAA	9720
GAGATGTCCA CTTGCAGGTA ATCCCAGAGT CTGCCCGGAT TGAGGGAGAT TATGCTAAGC	9780
TTTCTCGTAT CTTGGTGAAT CTGGTCGATA ACGCTTTTAA ATATTCTGCT CCAGGAACCA	9840
AGCTGGAAGT GGTGGCTAAG CTGGAGAAGG ACCAGCTTTC AATCAGTGTG ACCGATGAAG	9900
GGCAGGGTAT TGCCCCAGAG GATTTGGAAT ATATTTTCAA ACGCCTTTAT CGTGTCGAAA	9960
CTTCGCGTAA CATGAAGACA GGTGGTCATG GATTAGGACT TGCGATTGCG CGTGAATTGG	10020
CCCATCAATT GGGTGGGGAA ATCACAGTCA GCAGCCAGTA CGGTCTAGGA AGTACCTTTA	10080
CCCTCGTTCT CAACCTCTCT GGTAGTGAAA ATAAAGCCTA AAACCCCTTT ACAATCCAG	10140
CTATTCATGG TAGAATAGAT TTTGTGTGAA ATATCAGCAG GAAAGCATGA AGCTCGTCAA	10200
CAGGTGTCTT ATGACAAGTA ACCTTGGCTG TTTAGGCGAA GGCATCTGC ACGG	10254

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9769 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCGGCGACTA TCGATAACAC TTGACTTGGT AGCCCCACAT TTTGGACAAC GCATCCTTTC	60
CCTCCTTATC GTTTCTTTT CATTATACCA TTTTTAAGC GATTCCCAA ACAATCTTTC	120
TTTTTGCTTG ACAAGTTTT TGTTTGTTG TATTATTTAA TTAAGACAAC AAGTAAAAG	180
AAAGGAGACT AAGATGTCCT GGACATTGA CAACAAAAA CCCATCTATT TACAGATTAT	240
GGAGAAAATC AAGCTTCAGA TTGTTTCCCA TACTGGA CCAATCAAC AACTTCCAAC	300
CGTGAGGAGC TAGCTAGCGA GGCTGGTGT AATCCCAATA CCATCCAAAG AGCCTTATCA	360
GACCTTGAAC GAGAAGGATT TGTCTACAG AAGCGAACA CTGGACGATT TGTGACTAAG	420
GATAAGGAGC TAATCGCCCA GTCACGCAA CAATTATCAG AAGAAGAATT GGAACACTTC	480

319

GTTCCTCCA TGACCCATT TGGCTATGAA AAAGAAGAAC TACCAGGCGT AGTCAGTGAT	540
TATATTAAAG GAGTTTAAGC CTATGTCATT ACTAGTATTT GAAAATGTAT CCAAATCATA	600
TGGAGCAACA CCAGCCCTTG AAAATGTTTC TCTTGACATT CCAGCTGGAA AAATTGTCGG	660
CCTTCCTGGG CCAAACGGCT CAGGAAAAAC AACCTGATT AAATAATTA ATGGCCTCTT	720
ACAACCAGAT CAAGGACGTG TCCTCATCAA CGACATGGAC CCAAGCCCAG CAACCAAGGC	780
CGTTGTAGCT TATTTGCCTG ATACGACCTA TCTCAATGAG CAAATGAAGG TCAAAGAAGC	840
CCTAACCTAC TTCAAGACCT TCTATAAAGA TTGTCAGATC TTGAACGCGC CCATCATCTA	900
CTTGACAGACC TGGGCATTGA TGAAAATAGT CGTCTCAAGA AACTATCAAA AGGAAACAAA	960
GAAAAGGTTT AACTGATTTT GGTATGAGC CGTGATGCTC GTCTCTATGT TTTGGACGAA	1020
CCCATTGGTG GGGTGGATCC AGCAGCCCGT GCTTATATCC TCAATACCAT TATCAACAAC	1080
TACTCACCAA CTCTACCGT TTTGATTTCT ACCCACTTGA TTTCTGATAT CGAGCCAATC	1140
TTGGATGAAA TTGTCTTCCT AAAAGACGGA AAAGTCGTCC GTCAAGGAAA TGTAGATGAT	1200
ATTCGCTACG AGTCAGGTGA ATCCATTGAC CAACTCTTCC GTCAGaATT AAGGCCTAAG	1260
CAAAGGAGAT TATTTATGTT TTGGAATTTA GTTCGCTACG AATTTAAAAA TGTTAACAAG	1320
TGGTATTTAG CCCTCTACGC AGCCGTGCTA GTCCTTTCTG CCCTCATCGG AATACAGACA	1380
CAAGGCTTTA AAAATCTACC TTACCAAGAA AGTCAGGCTA CTATGCTACT TTTCTAGCT	1440
ACAGCTTTTG GTGGCTTGAT GCTTACACTT GGGATTTCAA CCATTTTCTT GATTATTAAA	1500
CGCTTCAAAG GTAGTGCTA CGACCGACAA GGCTATCTGA CTTTGACCTT GCCAGTTTCT	1560
GAACACCATA TCATCACAGC CAACTAATC GGTGCCTTTA TCTGGTCATT GATTAGCACC	1620
GCTGTATTGG CTCTAAGTGC TGTATTATT CTGGCTTTAA CAGCTCCAGA ATGGATTCCT	1680
CTTTCTTATG TGATTACATT TGTAGAAACA CATCTCCCTC AGATCTTTCT TACAGGTATA	1740
TCCTTCCTAC TAAATACTAT TTCAGGAATC CTCTGCATCT ACCTGGCTAT TTCCATTGGA	1800
CAGCTTTTCA ATGAATACCG TACAGCACTC GCTGTTGCAG TCTACATTGG TATCCAAATC	1860
GTCATTGGAT TTATTGAACT TTTCTTCAAT CTTAGTTCTA ATTTCTATGT CAATTCACCTG	1920
GTAGGACTCA ATGACCATT CTATATGGGA GCAGGTATAG CCATTGTTGA AGAACTCATA	1980
TTCATAGCTA TCTTTTATCT CGGAACCTAC TACATCTTGA GAAATAAGGT TAATTTGCTT	2040
TAAATAATTT TTACCTAGAT ATGTAACATA CTCATAGAAC AAAAGAGACC AGGCAAAAAG	2100
TCITTAAAAAT TAGAAAACGC ATAGTATCAG GTGTTGAATA TGTACTGCcC CCCAAAACTT	2160
AGATTTTTC TGTCTAACTT TGGGGGCAG TTCATAAGAA CCTTGGTAAT ATGCGTTTTT	2220

320

TGTGAGCTGA	CTTATTTCCT	TTCACCTATAT	CGCAAAATGA	AATAAGAACG	GAACGATGGG	2280
ATTTTGGAAT	TCAAATCAAT	TTATAAGAAT	GTTTTAGAAG	TAATATTATC	CTATTCCAGA	2340
TTCAGTTCAC	TATACAATTG	AGTTTTCAAG	CAACCTGTTT	ACATAATGTG	TACATAATTA	2400
GGTTCGTGAT	TCCACCCTTT	TCACCTTTAA	AAACCTCGCT	TTCGCAAGGC	TCTTCTATTT	2460
ATAAGATAAG	GCACGTTTAA	AGGTTTTCCA	AATCCCTAAA	TCATCCGTTT	GAAGAACGAG	2520
ACTAGCATAC	ATGCGTCCGA	TAAATCCTGT	TGCTACCACC	GCAAAAATCA	CTGTAATAGC	2580
AAGTGAAATC	CATGCTTCTG	CTCCCCCGC	ATAGTCATTA	ATCGTTCGAA	ACGGCATAAA	2640
GAAGGTCGAA	ATAAAGGGAA	TATAAGAACC	AATCTTCAAG	AGGAGATTGT	CACCAGCTGC	2700
ACCTAGAGCT	GTCACCTCAA	AAAAACCACC	CATAATCAAA	ATCATCAAAG	GCGACAAGGC	2760
TTTCCCTGAG	TCCTCAGGAC	GAGAAACCAT	AGATCCTAGG	AAGGCTGCCA	AGACTACGTA	2820
CATGAAAAGA	CTGATCAAAA	TAAAGAGCAA	GGTATTCAGT	GAGATAGCAT	CTCCCAAGTG	2880
ATCCAAAATA	CCAGACTGAG	CCAAGAATGG	CAAACTTTTA	AAGAGCAAAA	CGGCAGCCAG	2940
ACCACCTACA	ACATAGATCC	CAATATGCGT	TAAATCACT	AGAAACAGAG	CCATCATCCG	3000
CGCATAGAAA	TAGTGACTTG	CCCTTATGCT	AGAAAAACG	ACTTCCATAA	TTTTGGTGCC	3060
TTTTTCACTG	GCAACTTCCT	GAGCTGTAC	ACCCGCATAG	GTAATCAGAA	TCATATAAAG	3120
AAAGAATCCT	AAGGCACCTG	CTGCAATTGT	TTGAATAAAC	TTTTTATTTT	CCTTGCTTC	3180
ATCAATCTTT	TCTGTGAATT	GAATTGTCTG	CGCTAAGCGT	TTTTCCTGCT	CTTGAGACAA	3240
GGAAGCAGTT	GAACGATTAA	GCTGATTTTG	CAGTTCATTG	AGTGACCTG	TAACCTCAAA	3300
TTTAATTCCA	TTTTCAAGCG	ATGTTTCGCC	ATGATAAACT	GCCTTTAGAA	CACTATCTTC	3360
TTGATCAATG	GTCAAATAAC	CTTTTAATTT	TTCTTCTTTA	ATTGCTTCTT	TGGCACTTGC	3420
TTCTCTTTTA	TAGTCGAAGT	TAACACCATT	TACATCTTTC	AGTCCTTCTG	CTACAGATGG	3480
CACTGTTGTC	ACTACTGCCA	CTTTATTATT	TTTAGCCATA	GAAGAACCTT	GGAGATGCCC	3540
AATTCCTACA	GAGATTCCTA	AAAAGAGGAA	CGGCGAAATC	ACCATAAAGA	AGAAACTCCA	3600
TGACTCGACA	TGTCGAAGAT	AGGTTTCCTT	GATTACAACC	CACATATTTT	TCATACTTCC	3660
ACTCCTGATT	CTAGTTTAAA	GATTTTCATG	ATAGTTGGCG	CTTGTTGGTC	AAATGTTGCG	3720
ATATATTGAC	CTTGAGTCAA	GATTGAGAAG	AGTTCCTTTC	CAGCGCTCTC	ATCCTCCAAA	3780
ATCAATTPTC	AACTGCCTTG	TTTGGTCAAG	CTCACCTGTT	TGACATGAGG	AAGATTTTCC	3840
AATTCCTCCT	TGCTTCGTTT	ACTTGAAACA	AAGAGACGCG	TTTTCCCGTA	TTGATTGCGG	3900
ACATCCTGAA	CTGGTCCGTG	CAAGACCACA	CGGCCATCTC	GGATCATCAG	AATATCGTCA	3960
CAAAGTTCCT	CAACATTGGT	CATGACATGG	TCAGAAAAGA	TAATGGTTGT	CCGCGCTCTT	4020

321

TTTCCTGAAA AATGACTTGT TTGAGCAATT CTGTATTAAC TGGGTCCAAT CCACTAAAAG	4080
GCTCATCCAA GATAATCAGG TCTGGTTCAT GAATCAGAGT AATAATGAGC TGAATCTTCT	4140
GCTGATTTC TTTTGACAGA CTCTTGATTT TATCTGTCAG CTTTCCTTTC ACTTCCAACC	4200
TCTTCATCCA TTGAGGGAGT TTTTCITTGA CTCTTTTGGC ATCCATGCCT TTTAGAGTCG	4260
CCAAGTAGCG AACTTGTTCA AGAACTGTCA ATTTAGGCAT GAGATGCGTT CTTCAGGCAG	4320
ATAACCAATC CGAGCATAGG TCTCCTGACG AATATCCTGA CCATCCAGAC CGATTCTTCC	4380
CTGATATTCT AGGAATTTCA AAATACTATG GAAATCGTT GTTTTCCAG CACCATTTTT	4440
TCCGACTAGT CCCAAAATAC GACCTGGTCG CGCTTGAAAG TCAATACCAA ACAAACTTG	4500
CTTGGATCCA AAACCTTTCT CTAGACTTCT TACTTCTAGC ATCTTTCACC TCCGAAATTT	4560
CTTGCACTCA TTATACTCCT TTTTGATAGC CTTTACAATG TTTTGTGCC ATTTTGTAGAA	4620
GACTATTGCT GTGTAAAATA TGGCCTGGAG CACTTTTATA CTCAATGAAA ATCAAAGAGC	4680
AAACTAGGAA GCTAGCCGTA GACTGCTCAA AGTACAGCTT TGAGGTGCA GATAAACTG	4740
ACGAAGTCgA CTCAAAACAC TGTTTGGAG TTGTGGATAG AACTGACGAA kCrTAaCTAT	4800
ATCTACGGCA AGGCGAAcTG ACGTGGTTTG AAGAGATTTT CGAAGAGTAT TAGTGATAAA	4860
TCCATTATAC AGCAGCAAAC TTAATTTATA CCTCCGCTC CTCAACTGTC TATTTTAAAT	4920
CCTGAATTGT TATTTGAGTA ACTCCTTTTT CCTCGTAAAG TTTCTTCTCT CTAAACTTC	4980
TGGAAAAAGG CTAATAGTTT CAGACAACAT TTTTATAAGA AACAAGTCA TCTGTCATTT	5040
CAAGAAGGAG TAATCCTTTA TCTACTAATG GACGGAACAG AATTCAACCG CTGTCCGAT	5100
ATGTTTTCTA AGGATTATAT AGTAAATGA AATAAGAACA GGACAAATTG ATCAGGACAG	5160
TCAAATTGAT TTCTAACAAT GTTTTGAAG TAGATGTATA CTATTCTAGT TTCAATCTGC	5220
TATATCTATT ATGCACACCC CTATAGGATC TAATGAAAT CACAACAGGC TCATTCATAG	5280
ATGGTTACCT AAGCCTAAGG GAACTAAGAA AACGACTACC AAGGAAGTCG CATTCATCGA	5340
AAAGTAGATT AACAACTATC CTAAAAATG CTTGAACTAC AAGTCCCCCA GAGAAGACTT	5400
CTGGATGACT AACTTGAAC TGAATTTAG CAATAATTAA TTCACTATCT AACTATATTT	5460
AGTAATTATT TCAGAACTGA TTAATATTAA AATTAACTAA CAATTCAAAG GATTCATACT	5520
AGCCATAAAT TACGTCCATC AGAGAGAGAC TCTTACTACT TTTAGATTTT AGTCTTTCTA	5580
GCTTCAGAAT ACATCTAAAC TTTAGGGAAA ATGACTATTC GAAAGCGCGA ATGCCTCAA	5640
ATTATCTCAG ATAAGCTATT CGAACTTAG AATGCTTTTA AATTTATGGA ATTGCGATTA	5700
TTCGAAACCT AGAATGCATA TAACCTTTAG TTGACAGACC TATTCTAAGT CTCGAAGGGC	5760

322

TATTTACTTT CTATTCCTTA TCAAAAAAGA CTCATTCCCC CTTTCTCCTC CAAAATATGG	5820
TATAGTAGAA ATATACTATC TATGAGGAGT TTACATGTCA CAGGATAAAC AAATGAAAGC	5880
TGTTTCTCCC CTTCTGCAGC GAGTTATCAA TATCTCATCG ATTGTCGGTG GGGTTGGGAG	5940
TTTGATTTTC TGTATTGGG CTTATCAGGC TGGGATTTTA CAATCCAAGG AAACCCTCTC	6000
TGCCTTTATC CAGCAGGCAG GCATCTGGGG TCCACCTCTC TTTATCTTTT TACAGATTTT	6060
ACAGACTGTC GTCCCTATCA TTCCAGGGGC CTTGACCTCG GTGGCTGGGG TCTTTATCTA	6120
CGGGCACATC ATCGGACTA TCTACAACTA TATCGGCATC GTGATTGGCT GTGCCATTAT	6180
CTTTTATCTA GTGCGCTAT ACGGAGCTGC CTTTGTCCAG TCTGTCTGCA GCAAGCGCAC	6240
CTACGACAAG TACATCGACT GGCTAGATAA GGGCAATCGT TTTGACCGCT TCTTTATTTT	6300
TATGATGATT TGGCCCATTA GCCCAGCTGA CTTTCTCTGT ATGCTGGCTG CCCTGACCAA	6360
GATGAGCTTC AAGCGCTACA TGACCATCAT CATTCTGACC AAACCTTTTA CCCTCGTGGT	6420
TTATACCTAC GGTCTGACCT ATATTATTGA CTTTTTCTGG CAAATGCTTT GACACGTAAA	6480
AAATCCGTTT GGTTCCTCAA GTGGATTTT AAAGCGTAGA TTAATATAG CTTGATACTA	6540
AATATACTTT GGTATGGAAT TCATGCATAT TTTTCGATAG TGAGGCGAGG ACTTACCTAG	6600
CCTTTCCGCC GTGATAGAAA CACCTGAAAT CTAATGGTTT CAGGTATTCG GAAACTTTGA	6660
GCCTAGTGTC TCAAAGTTTA GGTATGGAAT TTTGAAGAAA GTCGCTACCG TCCGTAATCA	6720
CTTAAGGAAA GGCTCAAAAA TATTGTTTTT AACCACAAAA TCCGTTTGGT TTCCCAAGCG	6780
GATTTTGTGC TTTATTTTGA AACTTCTTTT GCAAGAACA AGTTCCCAAG TGTGGCAGAA	6840
CCATTTCTCTG CGACTGCTGG CGTCACGATA TAGTCACGCA CATCTGGTAC TGGTAGGTAA	6900
CCATTAAGAA GAGATGTAAA TTTCTCACGG ACACGGTCCA GCATATGTTG TTGAGCCATG	6960
ACCCCTCCAC CAAAGACAAT CACGTCTGGG CGGAAAGTCA CTGTCGCATT AACCGCAGCT	7020
TGAGCGATAT AGTAGGCTTG AACATCCCAA ACAGGGTTGT TGAGTTCAAT AGTTTCCCCA	7080
CGTACACCTG TACGAGCTTC CAACTTGGA CCAGCTGCAT AACCTTCTAG ACATCCCTTA	7140
TGGAAAGGAC AAACACCCTT AAACCTTTTT TCAATATCCA TTGGGTGTCT AGCAACATAA	7200
TAATGACCCA TTTCAGGGTG ACCCACACCA CCGATAAACT CACCACGTTG GATGACGCCT	7260
GCACCGATAC CTGTACCGAT TGTGTAGTAA ACCAAGTTTT CGATACGACC ACCAGCATTG	7320
TIACGGGCAA CCATTTTACC GTAAGCAGAG CTGTTTACGT CTGTTGTGAA GTACATTGGC	7380
ACGTTTAGGG CGCGACGAAG GGCACCAAGC AAGTCTACAT TTGCCAGTT TGGTTTGGGA	7440
GTGCTCGTGA TAAAGCCATA AGTTTTTGAG TTTTGTCAA TATCAATCGG CCCAAATGAA	7500
CCAACTGCAA GACCAGCAAG GTTATCGAAT TTTGAGAAGA ACTCAATGGT TTTATCGATT	7560

323

GTTCGATTG GAGTTGTTGT TGGAAATTGT GTTTTTTCTA CAACGTAAA GTTTCATCA	7620
CCGACAGCAC AGACAACTT TGTACCGCCC GCTTCCAAGC TTCCATATAA TTTTGTCTATG	7680
ATAAACCTCT TGTTTTATT TTCTTTATTA TAGCATACTT CGAAAGTCTA AATGTCTCTA	7740
TTTTTTAGAT TTTCTCTGT AAATCTTACT ATCTAATAAA AACGAACAAA CATGTCAATT	7800
GTTCGTTTTT ACATTAGAGA GGATTGATTA GATTTTCACT TCGATCACAG CATCCCCCTT	7860
AGCAACTGAA CCTGTTGCGA CTGGAGCTAC TGAAGCGTAG TCACCTGTAT TTGTAACGAT	7920
AACCATGTGT GTATCATCAA GTCCAGCTGC AGCGATTTTG TTTGAGTCAA ATGTTCCAAG	7980
AACATCGCCA GCTTTCACCT TATTACCTTG AGCAACTTTT GTTCAAAAC CGTCACCGTT	8040
CATAGATACA GTATCAATAC CAACATGAAT CAAAACCTCA GCACCATTTT TGTPTTCAA	8100
ACCAAAAGCG TGCCCTGTTG GAAAGGCAAT TGAAACTTCA GCATCAGCTG GTGCATAGAC	8160
CACGCCCTGG CTGGMTC AACCATACTT TGTCCCATTA GCTCCACTTG AGAAGACTGG	8220
GTCATTGACA TCAGCAAGAG CGACAACATC ACCGACGATA GGAGTTACAA GTGTTTCATT	8280
TTGAAGAGCT GCTGGCGCAA CTTCTTCTTT TTCTTCAGCC ACTTCAGCTC GTTTTGCGAC	8340
TGCAGTTGCG TCTACTTCAT CTTGTAACC AAACATGTAA GTAAGAGCAA AACCAGGGC	8400
AAATGATACA GCTACCATAA GAAGGTATTG TGGAAGTTGT CCGTTACCAA CATAAAGCAT	8460
TGTACCAGGG ATGATGGTGA TACCATTACC AGTACCAGCA AGTCCAAGGA TAGAAGCCAA	8520
TCCACCACCG ATGACACAG CAATCAATGA AAGGAAGAAT GGTTCACGGA AGCGCAAGTT	8580
CACCCCGAAG ATAGCAGGCT CTGTAATACC TAGGAAGGCA GAAAGAGCAG CCGGGAAAGC	8640
AAGTGTTTTC AGTTTGGAT TTTTGTGTTT AACACCAACC GCAACAGTAG CAGCACCTTG	8700
AGCTGTCATA GCAGCTGTGA TGATAGCGTT GAATGGGTTA GCATGGTCAG CAGCAAGTAA	8760
TTGCACTTCA AGCAAGTTGA AGATGTGGTG CACACCTGAC ACGACGATCA ATTGGTGAAC	8820
CCCACCAATC AAGAAACCAC CAAGACCAA TGGCATGCTA AGAATCGCTT TTGTAGCAAT	8880
AAGGATGTAG TTTTCAACAA CGTGGAAAC TGGTCCAATG ACAAAGAGTC CAAGGATAGA	8940
CATGACCAA AGTGTACGA ATGGTGTTAC CAAGAGGTCA ATGACATCTG GAACAACTTG	9000
CGGACAGCTT TTTCAAATTT AGCTCCGACA ACCCCGATGA TGAAGGCTGG AAGAACGGAA	9060
CCTTGCAAAC CAACAACAGG GATGAAACCA AAGAAGTTCA TCGCTGTTAC TTCACCACCT	9120
TGAGCAACTG CCCAAGCGTT TGGAAGTGAG CCAGAGACAA GCATCATACC AAGAACGATA	9180
CCAACGGCAG GATTTCCACC AAATACACGG AAGGTTGACC ACACAACCAA ACCTGGCAAG	9240
ATGATGAAGG CTGTATCTGT CAAGATTGT GTGTAAGTTG CAAAGTCACC TGGAAGTGGC	9300

324

ATTTCAAGAG CGTTGAAAAG ACCACGCACA CCCATGAAGA GACCTGTCGC TACGATAACT	9360
GGGATGATTG GAACGAAAAC ATCACC AAAA GTACGGATAG CACGTTGGAA CCAGTTCCT	9420
TGTTTAGCAA CTCTGCTTT CATGTCATCC TTAGATGATG TTGGTAATCC AAGTACAACA	9480
ACTTCATCGT ACATTTGTT AACTGTACCT GTACCAAAGA TAATTTGGTA TTGCCCTGAG	9540
TTAAAGAAAG CACCTTGAAC TTTTCCAAAG TTCTCAATCA CTCTTTTATT GATTTTCTCT	9600
TCATCTTTGA CCATGACACG TAGACGAGTC GCACAGTGGG CAACACTATT GACATTTTCA	9660
CGTCCGCCCA AGGCATCGAT GACTTTTTTT GCAATTTCTT GATTGTTTCAT TTGCAAAAAT	9720
CTCCTTATAT AACATTTTGT TCTGTTTGA AAGCGATTTT ATTCGCCGG	9769

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3149 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CGCTTGAGTG CTAATTCATA GTTCTATTGT ATCACTTGGT CAGAAATAAT CAAGAAAAAA	60
GTCTGACTTT CTCAAGATAA AAAGCCTGAG ACCAACTCAG ACTTTTAAAT TCTTAAATG	120
GCAATTCCTC CTCTTCCAAG ACCAAATCTG CCAAATCTTG GCCTGCATTA TTTTCACGCA	180
TAGCACGTTG GGCACGACTT TCCAAGAGTT GGAATCCTGT GACAAGTACT TCGGTCACGT	240
AGTTCAATTG GCCATTTTTC TCAAAGCGAC GGGTACGCAA TTCTCCATCA ACGGAAATGA	300
GACTACCTTT GGTGCGTAC TTGCCAAAGT TTCTGCTAGT CTGCCCCATA GGACCATATT	360
GACAAAATCA GCTTCACGTT CACCGTTTGT GTCTTTGTAA CGACGGTCA CAGCGATAGT	420
TGCTCGCGCT ACCGACTTGT CATTTGTTGGT TTTGTGCAAT TCTGGTGTAG ACGTTAAACG	480
TCCAATCAAG ATAACCTTAT TATACATATT TTCTTCCTCC TACTTATCTA TTCGTAGGAA	540
ATCAAAAAAA GTTACAGAAA TTTGTAACCT TFCGAGAAAA TTTTATTATT TTTATGAACC	600
ATGAAACCTG TCGCCTGTTG ATTGGCCATA ATGGTCATAT CTGTAATCTG AACACGACGA	660
GGTTGACTAG TCACATAGAC TACTGTATCT GCAATATCCT GAGCTTGCAA AGCTTCTATT	720
CCTTGGTAAA CGGACGCAGC TCGTTCCTTA TCACCATGAA AACGCACTGT AGAAAAATCT	780
GTTCGACAA TTCCAGGCTG AATGGTCGTC ACCTTGATAT CCGTTGCGAT GGTATCAATT	840
CGCAGTCCAT CTGAAAAGGT CTTAACTGCC GCCTTGGTGG CTGAGTAAAC AGCTGCACCA	900
GCATAGGCAT AAATTCCTGC GGTGACCCC ATATTGATAA TATGACCTTG ATTGGCTTTT	960

325

ACCATTGCTG GCAAGAAACA GCGAGTGACT GCCATCAAAC CTTTGACATT GGTATCCAAC	1020
ATGGTCAGCA TATCCAAC TCATAGTCT TGATAGGGAG CTAAGCCAAG AGCCAGTCCT	1080
GCGTTATGA CCAGGATGTC AATCTGACCT ATCGTTTCTA AAATATCAGA GCAGACAGTC	1140
TTTACCATTG TCATATCCGT GACATCTAGG AGAAAAGTCC AAAGTGTG ATTTGGAAAA	1200
GTTTCTGCAA ACTCCGCCTT AAGAGCTTCT AGTCTGTCTA TCCGTCGTCC TGTTAGAACG	1260
ACATCCTCAC CCTGCTCCAG ATAAGCACGC GCAATCGCTT CACCGATTCC TGATGTCGCT	1320
CCTGTAATCA CAACATTTTT TGCCATCTTA TTTCCTTCTA GCTGGTCTAT CAGATATTAA	1380
CAACTTCTTA GGCAGTCCAG TGTTCGCTG GGTGCAACGG TGTTCGACA ACTTGGTCTT	1440
CTGATAATTC AAGCACCCCA CGTTTTGTG GAGCATTTGG CAGATGCAAT TCACGAGGAC	1500
TGCACATCAT ACCAAAATC TTTCACCAC GAAGTTCACC TGGGAAAATG AGATTCCCTT	1560
TTGGCATCAT AGCTCCAGGA AGCGCGACAA TGGTTTTCAA CCCACACGC GCATTGGGAG	1620
CTCCTGCAAC GATTGTGACA GTCTTATCAC TTGCGACTGC AACTTGGCAG ATGTTGAGGT	1680
GGTCACTATC TGGATGGGCT ACCATCTCAA CAATTTTACC TACAACAAAC TTAGGTTCTT	1740
TATCATTAAC AATTCTTCT GTAAAACCTT CCGCTGCAA CTCTTGGTTC AAACGAGCGA	1800
CTTGCTCATC TGTCAAAAAG ACTTGACCGC GCTCTGCAAT TTCAAATAAA CTTGAAACTT	1860
CGAAAATATT CCAAGCCACT GTTCCCAT TATCTTTGAG AAAACACGG GCTACCTTGC	1920
CTTTGCGCTC CACATCCAGT TTGGCATCTC CGTATTTTT CACGATGACC ATAAGGACAT	1980
CACCGACATG TTCTTTATTA TATGTAAAA TCATGTGTTT CTTTTCTCC TATTTAGTC	2040
CTGCTAAAAA GTCATTGATT TGTGCTGTC TTTTACGGTC GCGATTGACA AAACGACCGA	2100
TTTCCTTGTC CTTTCTAGA ACAACAAGGC TAGGAATTCC GTAAACATCC CAGAGTTTGG	2160
CCAAATCCAT ATACTGATCT CGGTCCATTC GAATAAAGGT GAACTCTGGA TTGGTCTCCT	2220
CAATCTCTGG TAAGGCAGGA TAAATATAAC GACAATCGCT ACACCAGTCT GCCACAAAAA	2280
TGAAGACCTT CTGCCCCGCT TTTTCCACTA AAGATGCTAA TTCTTCTAAA CTTGCTGGCT	2340
GTATCATAAG ACTTCCTCCT CATAGACTAG GTCTTCATTT TCATAGACAA AGGTATAATG	2400
ACGGCCATCC TCAAAAATGA CGCCACCAAC CAAGCTCTCC AGACTGCTTT CGTAAACTTG	2460
AACATAAAGG GTCGCAATTT CCCCCATGTC GGAAAAATGG TCTCGCACAA TCTCTGTCAA	2520
CTCTTCCTGA GTCTTCATGA GCTTACGGTC ATCTGCAACT TTTTTCGTAG CAAGAGCAAG	2580
GCTTCCGATA CCTAGCAGAG CCAAGCCTGC CATCCACATT TTTTGTAGCTT TCATACCATT	2640
CATTTTAACA CAAAAAAGGC TTCAGGACAA ATGAGGAAGC AGCAGAAAAG CAAGTAAAAA	2700

326

GCCTCTTCCT TTAAGGAAAA GGACTTCTTA TACTCAATGA AAATCAAAGA CCAAAGTAGG	2760
AAGCTAGCCG CAGGCTGCTC AAAGCACTGC TTTGAGGTTG TAGATAGAAC TGACGAgTca	2820
CTCAAAACAC TGT'TTTGAGG TTGTGGATGA AGCTGACGTG GTTTGAAGAG ATTTTCGAAG	2880
AGTATTATTC TTATTGCCAG GCACCTAAGT TGCCAACGTA GTAACATATCA GGTGTGTAGG	2940
TATTGCGAGC ATCTTACCTG ATGAAGCCAG ATAATACTAC TTGCCATTGT CTTTGACCCA	3000
ATCATTCGCA ATCATGGAAC CAGAAGAACT TACATAATAC CATTCTCCCT TGTCTATAAC	3060
CCAAGTACTG ACTTTCATGG TTCCTGAGCA ATTAAAGGCA AAAAACTGT CCAATAACAT	3120
TCGTTTTTTA AAAGCATTTG ACACTACAT	3149

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10240 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CCAAAAATTC AACCTTTAAG GGGAGTCCAG AGAGACTCAC AAGGTGTCAG ATAAAAGAAT	60
GGTGCAATTT TCTAGAGGAG ACTTTTGTAG TGTGCTCTCT TGTGTTGTAC GATTTTAACT	120
GAGGCCTTGC ACTAGCAAGG TCTTTCTTTT ATCTGGTCCC CTTAAAATTT AAGGAGGAAA	180
AGTTATGAAT CCCACATGTA AGAAGCGTTT GGGTGTCAAT CGGTGGAATA CCATGAAGGT	240
GGTTGCACAA GAGGAAATCG CGCCACAATC TTTGAATTAG TCCTAGAAGG AGAAATGGTT	300
GAAGCCATGC GAGCAGGCCA ATTTCTTCAT CTGCGTGTAC CGGACGATGC CCATCTCTTA	360
CGTCGTCCTA TTTCAATTTC GTCTATTGAC AAGGCAAACA AGCAGTGTCA CCTCATTTAT	420
CGGATTGACG GAGCTGGGAC TGCAATTTT TCAACCTTAA GTCAGGGAGA CACTCTTGAT	480
GTGATGGGGC CTCAGGGAAA TGGT'TTTGAC TTGTCTGACC TTGATGAGCA GAATCAGGTT	540
CTCCTTGTG GTGGTGGGAT TGGTGTTCCT CCCTTGCTTG AGGTGGCCAA GGAATTCAT	600
GAACGTGGAG TGAAGTAGT GACAGTCCTC GGTTTGTCTA ATAAGGATGC TGTATTTTGT	660
AAAACGGAAT TGGCTCAGTA TGGTCAGGTC TTTGTAACGA CAGATGATGG TTCTTATGGC	720
ATCAAGGGAA ATGTTTCCGT TGTATCAAT GATTTAGACA GTCAGTTTGA TGCTGTTTAC	780
TCGTGTGGGG CTCCAGGAAT GATGAAGTAT ATCAATCAA CTTTGTATGA TCACCAAGA	840
GCCTATTTAT CTCTGGAATC TCGTATGGCT TGTGGGATGG GAGCTTGCTA TGCCTGTGTT	900
CTAAAAGTAC CAGAAAACGA GACGGTCAGC CAACGCGTCT GTGAAGATGG TCCTGTTTTC	960

327

CGCACAGGAA CAGTTGTATT ATAAGGAGAA AATTATGACT ACAAATCGAT TACAAGTTTC	1020
TCTACCTGGT TTGGATTTGA AAAATCCGAT TATTCCAGCA TCAGGCTCTT TTGGCTTTGG	1080
ACAAGAGTAT GCCAAGTACT ATGATTTAGA CCTTTTAGGT TCTATTATGA TCAAGGCGAC	1140
AACCCTTGAA CCACGTTTTG GGAATCCAAC TCCAAGAGTG GCAGAGACGC CTGCTGGTAT	1200
GCTCAATGCA ATTGGCTTGC AAAATCCTGG TTTAGAGGTT GTTTTGGCTG AAAAGCTACC	1260
TTGGCTGGAA AGAGAATATC CAAATCTTCC TATTATTGCC AATGTAGCTG GTTTTTCAAA	1320
ACAAGAGTAT GCAGCTGTTT CTCATGGGAT TTCCAAGGCA ACTAATGTAA AAGCTATCGA	1380
GCTCAATATT TCTTGTCCTA ATGTTGACCA CTGTAATCAT GGACTTTTGA TTGGTCAAGA	1440
TCCAGATTTG GCTTATGATG TGGTGAAAGC AGCTGTGGAA GCCTCAGAAG TGCCAGTTTA	1500
TGTCAAATTA ACCCCGAGTG TGACCGATAT CGTTACTGTC GCAAAGCTG CAGAAGATGC	1560
GGGAGCAAGT GGCTTGACCA TGATCAATAC TCTGGTTGGA ATGCGCTTTG ACCTCAAAAC	1620
TAGAAAACCA ATCTTGGCCA ATGGAACAGG TGGAATGTCT GGTCCAGCAG TCTTTCCAGT	1680
AGCCCTCAAA CTCATCCGCC AAGTTGCCCA AACAACAGAC CTGCCTATCA TTGGAATGGG	1740
AGGAGTGGAT TCGGCTGAAG CTGCCCTAGA AATGTATCTG GCTGGGGCAT CTGCTATCGG	1800
AGTTGGAACA GCTAACTTTA CCAATCCTTA TGCTGCCCT GACATCATCG AAAATTTACC	1860
AAAAGTCATG GATAAATACG GTATTAGCAG TCTGGAAGAA TCCCGTCAGG AAGTAAAGA	1920
GTCTCTGAGG TAAACTGCAA TCAATCTGTT CTTGATTTTT TATTAGTTTG TAATATGAAT	1980
TTAGGAGAAT TTTGGTACAA TAAAATAAAT AAGAACAGAG GAAGAAGGTT AATGAAGAAA	2040
GTAAGATTTA TTTTITTAGC TCTGCTATTT TTCTTAGCTA GTCCAGAGGG TGCAATGGCT	2100
AGTGATGGTA CTTGGCAAGG AAAACAGTAT CTGAAAGAAG ATGGCAGTCA AGCAGCAAAT	2160
GAGTGGGTTT TTGATACTCA TTATCAATCT TGGTTCTATA TAAAAGCAGA TGCTAACTAT	2220
GCTGAAAATG AATGGCTAAA GCAAGGTGAC GACTATTTTT ACCTCAAATC TGGTGGCTAT	2280
ATGGCCAAAT CAGAATGGGT AGAAGACAAG GGAGCCTTTT ATTATCTTGA CCAAGATGGA	2340
AAGATGAAAA GAAATGCTTG GGTAGGAACT TCCTATGTTG GTGCAACAGG TGCCAAAGTA	2400
ATAGAAGACT GGGTCTATGA TTCTCAATAC GATGCTTGGT TTTATATCAA AGCAGATGGA	2460
CAGCACGCAG AGAAAGAATG GCTCCAAATT AAAGGGAAGG ACTATTATTT CAAATCCGGT	2520
GGTTATCTAC TGACAAGTCA GTGGATTAAT CAACCTTATG TGAATGCTAG TGGTGCCAAA	2580
GTACAGCAAG GTTGGCTTTT TGACAAACAA TACCAATCTT GGTTTTACAT CAAAGAAAAT	2640
GGAAACTATG CTGATAAAGA ATGGATTTTC GAGAATGGTC ACTATTATTA TCTAAAATCC	2700

328

GGTGGYTACA TGGCAGCCAA TGAATGGATT TGGGATAAGG AATCTTGGTT TTATCTCAAA	2760
TyTGATGGGA AAATtGCTGA AAAAGAATGG GTCTACGATT CTCATAGTCA AGCTTGGTAC	2820
TACTTCAAAT CCGGTGGTTA CATGACAGCC AATGAATGGA TTTGGGATAA GGAATCTTGG	2880
TTTTACCTCA AATCTGATGG GAAAATAGCT GAAAAAGAAT GGGTCTACGA TTCTCATAGT	2940
CAAGCTTGGT ACTACTTCAA ATCTGGTGGC TACATGGCGA AAAATGAGAC AGTAGATGGT	3000
TATCAGCTTG GAAGCGATGG TAAATGGCTT GGAGGAAAA CTACAAATGA AAATGCTGCT	3060
TACTATCAAG TAGTGCCTGT TACAGCCAAT GTTTATGATT CAGATGGTGA AAAGCTTTCC	3120
TATATATCGC AAGGTAGTGT CGTATGGCTA GATAAGGATA GAAAAAGTGA TGACAAGCGC	3180
TTGGCTATTA CTATTTCTGG TTTGTCAGGC TATATGAAAA CAGAAGATT ACAAGCGCTA	3240
GATGCTAGTA AGGACTTTAT CCCTTATTAT GAGAGTGATG GCCACCGTTT TTATCACTAT	3300
GTGGCTCAGA ATGCTAGTAT CCCAGTAGCT TCTCATCTTT CTGATATGGA AGTAGGCAAG	3360
AAATATTATT CGGCAGATGG CCTGCATTTT GATGGTTTGA AGCTTGAGAA TCCCTTCCTT	3420
TTCAAAGATT TAACAGAGGC TACAACTAC AGTGCTGAAG AATTGGATAA GGTATTAGT	3480
TTGCTAAACA TTAACAATAG CCTTTTGAG AACAAGGCG CTACTTTTAA GGAAGCCGAA	3540
GAACATTACC ATATCAATGC TCTTTATCTC CTGCCCATA GTGCCCTAGA AAGTAACTGG	3600
GGAAGAAGTA AAATTGCCAA AGATAAGAAT AATTTCTTTG GCATTACAGC CTATGATACG	3660
ACCCCTTACC TTTCTGTAA GACATTTGAT GATGTGGATA AGGGAATTTT AGGTGCAACC	3720
AAGTGGATTA AGGAAAATTA TATCGATAGG GGAAGAACTT TCCTTGGAAG CAAGGCTTCT	3780
GGTATGAATG TGAATATGC TTCAGACCCT TATTGGGGCG AAAAAATGC TAGTGTGATG	3840
ATGAAAATCA ATGAGAAGCT AGGTGGCAAA GATTAGTACT ATAAGTGAAT ATGATTTGAG	3900
TGAATAGTAA GTTAAAAATC CTGATTTCAA GTAAAATCAG GATTTTTTCA TGGATGCAAT	3960
TTTTTTGGAG TCTGGTGTGA CGCGGAGGGT CTTTTGTCCT GTGTAAGTGA CAAAGCCGGG	4020
TTTTCCACCA GTTGGTTTAT TGAGTTTCTT GACTTCAATC ATATCTACCT GCACCAGATT	4080
CGACAGGCGC CCTTGAGAGA AGTAGGCAGC TAACTCTGCT GCGTCTGTCT TGA CTGCATC	4140
AGATGGGTCA AGATTTCCCTG AGATGACAAC ATGGCTTCCA GGAATGTCCT TAGCATGGAA	4200
CCAAAGTTCC TCCTTGCGGG CCATTTTAAA GGTCAATTCC TCATTTTGAA GATTGTTTCG	4260
TCCGACATAG ATGATGGTTT TGCCATCGCT TGCTAGATAT TGTCTAGTT TTTTGCGTTT	4320
CTGGATTTTC TCCCGTTGTC TTCTGCGGAT AAAACCTGTT TGAATCAATT CTTACGGAT	4380
TTCAGCGATT TCTTCCAGTC CAGCTTGGTT GAGGACGGTT TCTACACTTT CCAGATAGAG	4440
AATAGTGGCT TTGGTTTCTT CAATCAAATC AGTCAAGTAT TTGACAGCTT CTTTGAGTTT	4500

CTGATACCGT TTAAATAGC GTTGGGCATT CTGGTTGGGA GTCAGAGCCT TATCAAGCGC	4560
AATCATGATA GGTGGTTGG TATAGTAGTT GTCTAGGATA ACCTGGTCTT GGTCGTTAGG	4620
CACTTGGTGG AGGAAGGTTG TCAGCAATTC TCCTTTTGA CGAAATTCCT CAGCGTTGTC	4680
TGTCGCCAGT AACTCTTTT CCTGTTTTT GAGTTTGTGT CGGTTTTCT GAAGTTCATT	4740
TTCAACACGA CGAATCAGTT CACTGGCCTG CTGTTTGACG CGGTCGCGCT CAGCCTTATC	4800
CTTATAGTAG GTGTCCAACA AATCAGAAAG ATTTGCAAAA GGCTCTCCCA CCTGATTGTC	4860
AAAAGGAAGT GGACTGAAGG AAGTCTCAGT CAAGCATGGC TTGGTTTCTT GATTGAAAAA	4920
ATTTTCGAAA GCGGAAAGTT TTCACTAAC CAGTATCCTT TCCAATTCAT TGCCGTATC	4980
GCGTCCCAGA CCTTGAAAGA GGCTTTGAAG ATTTTGTGCT GTTAGTTCTT GGGTTGTCAG	5040
GATTTCAAAG AGCTTTTCAT CCTTGATAGT AAAAGGATTG AGAGATTGTC TACTTGCGCG	5100
AGCGATATAG GTCGATCCTG GAAGTAAGGT GCGGTAGCTA TTTTGTGAAA AGCCGACGTG	5160
TTTGATAACT TCGAGGATTT TATGACTGCT TTTATCGACC AGTAGAATAT TACTGTGTTT	5220
CCCCATAATT TCGATAATCA AGGTAGCCTG GATATGGTCT CCAATCTCGT TTTTATTGGA	5280
AACTGTAATT TCCACAATAC GGTCAATTTT CACTTGCTCA ATCGACTCAA TCAGGGCCCC	5340
CTGCAAAATC TTTCTCAAAA CCATGATAAA GGTAGAAGGT TGAGCTGGAT TTTCAAAGT	5400
CGTTTGGGTC AGCTGAATGC GTCCAAAAC TGGATGGGCA GAAAGGAGCA GGCGATGGCT	5460
TTGGCGATTG CTGCGGATTT GCAAGACCAA CTCTGTGTC AAGGCTGAT TGATTTCTG	5520
GATGCGACCA TTCACTAATT CGCTTCGCAA TTCTCAACT ATGTGGTGTA AAAAAATCC	5580
GTCAAATGAC ATCGTTCTCT CCTGTGATT GTATTCCATA GTATTATATC AAAAGGTAG	5640
AATAAAATCA TGGAAATGTG GTATAATAAA GCCAAGTAAA GAGAAACGAG AAGCACATGT	5700
ATATTGAAAT GGTAGATGAA ACTGGTCAAG TTCAAAAGA AATGTTGCAA CAAACCAAG	5760
AAATTTTGA ATTTGCAGCC CAAAAATTAG GAAAGAAGA CAAGGAGATG GCAGTCACTT	5820
TTGTGACCAA TGAGCGTAGT CATGAACTTA ATCTGGAGTA CCGTAACACC GACCGTCCGA	5880
CAGATGTCAT CAGCCTTGAG TATAAACAG AATTGGAAAT TGCCTTTGAC GAAGAGGATT	5940
TGCTTGAAAA TTCAGAATG GCAGAGATGA TGTCTGAGTT TGATGCCTAT ATTGGGAAT	6000
TGTTTATCTC TATCGATAAG GTCATGAGC AGGCCGAAGA ATATGGTCAC AGCTTTGAGC	6060
GTGAGATGGG CTTCTTGGA GTACACGGCT TTTTACATAT TAACGGCTAT GATCACTACA	6120
CTCCGGAAGA AGAAGCGGAG ATGTTGGGTT TACAAGAAGA AATTTTGACA GCCTATGGAC	6180
TCACAAGACA ATAAACGAAA ATGGAAAAAT CGTGACTTGA TATCCAGTTT AGAATTTGCT	6240

330

TTGACAGGTA TTTTACTGTC TATCAAGGAA GAACGCAATA TGCGAAAACA CGCAGTGACG	6300
GCTCTAGTGG TCATCCTTGC AGGTTTTGTT TTTCAGGTGT CACGAATCGA ATGGCTCTTT	6360
CTCCTATTGA GTATTTTCTT GGTAGTAGCC TTTGAGATTA TCAACTCTGC TATTGAAAAT	6420
GTGGTGGATT TGGCCAGTCA CTATCACTTT TCCATGCTGG CTAAAAATGC CAAGGATATG	6480
GCGGCCGCGC CGGTATTAGT GGTTCCTCTT TTCGCAGCCT TAACAGGCGC ATTGATTTTT	6540
CTCCCACGAA TCTGGGATTT ATTATTTTAA ACAGTAAGAG GAAATTATGA CTTTPTAAATC	6600
AGGCTTTGTA GCCATTTTAG GACGTCCCAA TGTGGGAAG TCAACCTTTT TAAATCACGT	6660
TATGGGGCAA AAGATTGCCA TCATGAGTGA CAAGGCGCAG ACAACGCGCA ATAAATCAT	6720
GGGAATTTAC ACGACTGATA AGGAGCAAAT TGTCTTTATC GACACACCAG GGATTCACAA	6780
GCCTAAAACA GCTCTCGGAG ATTTTCATGGT TGAGTCTGCC TACAGTACCC TTCGCGAAGT	6840
GGACACTGTT CTTTTCATGG TGCCTGCTGA TGAAGCGCGT GGTAAGGGGG ACGATATGAT	6900
TATCGAGCGT CTCAAGGCTG CCAAGGTCC TGTGATTTTG GTGGTGAATA AAATCGATAA	6960
GGTCCATCCA GACCAGCTCT TGTCTCAGAT TGATGACTTC CGTAATCAAA TGGACTTTAA	7020
GGAAATTGTT CCAATCTCAG CCCTTCAGGG AAATAACGTG TCTCGTCTAG TGGATATTTT	7080
GAGTGAAAAT CTGGATGAAG GTTCCAATA TTTCCCGTCT GATCAAATCA CAGACCATCC	7140
AGAACGTTTC TTGGTTTCAG AAATGGTTCG CGAGAAAGTC TTGCACCTAA CTCGTGAAGA	7200
GATTCGCGAT TCTGTAGCAG TAGTTGTGA CTCTATGAAA CGAGACGAAG AGACAGACAA	7260
GGTTCACATC CGTGCAACCA TCATGGTCTGA GCGCGATAGC CAAAAGGGA TTATCATCGG	7320
TAAAGGTGGC GCTATGCTTA AGAAAAATCGG TAGCATGGCC CGTCGTGATA TCGAACTCAT	7380
GCTAGGAGAC AAGGTCTTCC TAGAAACCTG GGTCAAGGTC AAGAAAAACT GGCGCGATAA	7440
AAAGCTAGAT TTGGCTGACT TTGGCTATAA TGAAAGAGAA TACTAAGTAG AGGTAGGCTC	7500
ATGCCTGCTT CTTGTTTTTA CAGAAGGAGG ACTTATGCCT GAATTACCTG AGGTTGAAAC	7560
CGTTGTGCTT GGCTTAGAAA AATTGATTAT AGGAAAGAAG ATTTTCGAGTA TAGAAATTCG	7620
CTACCCCAAG ATGATTAAGA CGGATTTGGA AGAGTTTCAA AGGGAATTGC CTAGTCAGAT	7680
TATCGAGTCA ATGGGACGTC GTGGAAAATA TTTGCTTTTT TATCTGACAG ACAAGGTCTT	7740
GATTTCCCAT TTGCGGATGG AGGGCAAGTA TTTTACTAT CCAGACCAAG GACCTGAACG	7800
CAAGCATGCC CATGTTTTCT TTCATTTTGA AGATGGTGGC ACGCTTGTTC ATGAGGATGT	7860
TCGCAAGTTT GGAACCATGG AACTCTTGGT GCCTGACCTT TTAGACGTCT ACTTTATTTT	7920
TAAAAAATTA GGTCTTGAAC CAAGCGAACA AGACTTTGAT TTACAGGTCT TTCAATCTGC	7980
CCTTGCCAAG TCCAAAAAGC CTATCAAATC CCATCTCCTA GACCAGACCT TGGTAGCTGG	8040

331

ACTTGGCAAT ATCTATGTGG ATGAGGTTCT CTGGCGAGCT CAGGTTTCATC CAGCTAGACC	8100
TTCCCAGACT TTGACAGCAG AAGAAGCGAC TGCCATTTCAT GACCAGACCA TTGCTGTTTT	8160
GGGCCAGGCT GTTGAAAAAG GTGGCTCCAC CATTCCGACT TATACCAATG CCTTTGGGGA	8220
AGATGGAAGC ATGCAGGACT TTCATCAGGT CTATGATAAG ACTGGTCAAG AATGTGTACG	8280
CTGTGGTACC ATCATTGAGA AAATTCAACT AGGCGGACGT GGAACCCACT TTTGTCCAAA	8340
CTGTCAAAGG AGGGACTGAT GGGAAAAATC ATCGGAATCA CTGGGGGAAT TGCCTCTGGT	8400
AAGTCAACTG TGACAAATTT TCTAAGACAG CAAGGCTTTC AAGTAGTGA TGCCGACGCA	8460
GTCGTCCACC AACTACAGAA ACCTGGTGGT CGTCTGTTTG AGGCTCTAGT ACAGCACTTT	8520
GGGCAAGAAA TCATTCTTGA AAACGGAGAA CTCAATCGCC CTCTCCTAGC TAGTCTCATC	8580
TTTTCAAATC CTGATGAACG AGAATGGTCT AAGCAAATTC AAGGGGAGAT TATCCGTGAG	8640
GAAGTGGCTA CTTTGAGAGA ACAGTTGGCT CAGACAGAAG AGATTTTCTT CATGGATATT	8700
CCCCTACTTT TTGAGCAGGA CTACAGCGAT TGGTTTGCTG AGACTTGGT GGTCTATGTG	8760
GACCGAGATG CCCAAGTGGA ACGCTTAATG AAAAGGGACC AGTTGTCCAA AGATGAAGCT	8820
GAGTCTCGTC TGGCAGCCCA GTGGCCTTTA GAAAAAAGA AAGATTTGGC CAGCCAGGTT	8880
CTTGATAATA ATGGCAATCA GAACCAGCTT CTTAATCAAG TGCATATCCT TCTTGAGGGA	8940
GGTAGGCAAG ATGACAGAGA TTAAGTGGAA GGATAATCTG CGCATTGCCT GGTGTTGGTAA	9000
TTTTCTGACA GGAGCCAGTA TTTCTTTGGT TGTACCTTTT ATGCCCATCT TCGTGAAAAA	9060
TCTAGGTGTA GGGAGTCAGC AAGTCGCTTT TTATGCAGGC TTAGCAATTT CTGTCTCTGC	9120
TATTTCCGCG GCGCTCTTTT CTCCTATTG GGTATTCTT GCTGACAAAT ACGGCCGAAA	9180
ACCCATGATG ATTCGGGCAG GTCTTGCTAT GACTATCACT ATGGGAGGCT TGGCCTTTGT	9240
CCCAAATATC TATTGGTTAA TCTTTCTTCG TTTACTAAAC GGTGTATTG CAGGTTTTGT	9300
TCCTAATGCA ACGGCACTGA TAGCCAGTCA GGTTCCAAAG GAGAAATCAG GCTCTGCCTT	9360
AGGTACTTTG TCTACAGGCG TAGTTGCAGG TACTCTAACT GGTCCCTTTA TTGGTGGCTT	9420
TATCGCAGAA TTATTGGCA TTCGTACAGT TTTCTTACTG GTTGGTAGTT TTCTATTTTT	9480
AGCTGCTATT TTGACTATTT GCTTTATCAA GGAAGATTTT CAACCAGTAG CCAAGGAAAA	9540
GGCTATTCCA ACAAAGGAAT TATTTACCTC GGTAAATAT CCCTATCTTT TGCTCAATCT	9600
CTTTTTAACC AGTTTGTCA TCCAATTTTC AGCTCAATCG ATTGGCCCTA TTTTGGCTCT	9660
TTATGTACGC GACTTAGGCG AGACAGAGAA TCTTCTTTTT GTCTCTGGTT TGATTGTGTC	9720
CAGTATGGGC TTTTCCAGCA TGATGAGTGC AGGAGTCATG GGCAAGCTAG GTGACAAGGT	9780

332

GGGCAATCAT CGTCTCTTGG TTGTCGCCCA GTTTTATTCA GTCATCATCT ATCTCCTCTG	9840
TGCCAATGCC TCTAGCCCCC TTCAACTAGG ACTCTATCGT TTCTCTTTTG GATTGGGAAC	9900
CGGTGCCTTG ATTCCCGGGG TTAATGCCCT ACTCAGCAA ATGACTCCCA AAGCCGGCAT	9960
TTTCAGGGTC TTTGCCTTCA ATCAGGTATT CTTTATCTG GGAGGTGTTG TTGGTCCCAT	10020
GGCAGGTCT GCAGTAGCAG GTCAATTGG CTACCATGCT GTCTTTTATG CGACAAGCCT	10080
TTGTGTGCC TTTAGTTGTC TCTTTAACCT GATTCAATTT CGAACATTAT TAAAAGTAAA	10140
GGAAATCTAG TGCAGATAA AATCAATCTC AAATGCTCCT CTTGTGGCAG TATCAATTAC	10200
CTAACCAGTA AAAATTCAA AACCCATCCA GACAgATTGA	10240

(2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13206 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCTTTATCG TGGACGTGGT CAAGCCGAGA ATTCATCAA GGAGATGAAG GAGGGATTTT	60
TTGGCGATAA AACGGATAGT TCAACCTTAA TCAAAAACGA AGTTCGTATG ATGATGAGCT	120
GTATCGCCTA CAATCTCTAT CTTTCTCTCA AACATCTAGC TGGAGGTGAC TTCCAACTT	180
TAACAATCAA ACGCTTCCGC CATCTTTTTC TTCACGTGGT GGGAAAATGT GTTCGAACAG	240
GACGCAAGCA GCTCCTCAA TTGTCTAGTC TCTATGCCTA TTCCGAATTG TTTTCAGCAC	300
TTTATCTTAG GATTAGAAAA GTCAACCTGA ATCTTCCTGT TCCTTATGAA CCACCTAGAA	360
GAAAAGCGTC GTTAATGATG CATTAAAGAA CAGTCGAGAT GAAAAAATCG TGTGACGCAC	420
CAAGGGAGGA GTCTGCCCTT TTGAGGAAAT CTAGCGAGGA AAAACGATAC TGGAACAGCA	480
GAAAGTAAAA CTGACCTCAT GAGGAGGAAG AAAGTGGCTC ATGAGGTCAG GGGTTTGTAA	540
AGTTACATCT AGTTGAGAGA GGTATGAATG ATTTGGGATT AATCATTTCT TGTTTTAAAT	600
CAGGAGAATA GTAACGATTT TTCTCTTTT TGACGAATC TATTCCGTAA CGATCAATCA	660
ATTTAATCAT GTACCTAATA TTAGAATTGT TTATCCCAA TTTATTTGAA AGCTTCTCTA	720
AGCTATATCC TTGTTTTCTA AGTTCATAGA TCTGAATTT ATCATCATAA GTTAGTTTCA	780
TAATAAAAAC ACCCCAAAAG TTAGATTTTT TCTGTCTAAC TTTTGGGGGG CAGTTCATTC	840
AACACCTGAT ACTATGCGTT TTTCTTATTT GAAATACTTT TTAACAACC TCTTTATACT	900
CAATGAAAAT CAAAGTGCAA ACTAGAAAGC TAGCCTCAGG CTGCTCAAAA CAGTGTTTTG	960

333

AGGTTGCAGA TGGAAGCTGA CGTGGTTTGA AGAGATTTTC GAAGAGTATT ACTTAATCTT	1020
CTTGATACTT TGAATAAGAA TAAATCCTAC AATCATCCCT ACCATATTTT GCATAAAATT	1080
CGGTAGAATT TCTGGGAGGG CTGCTGCCCA GCCATTTCATC AAAGCAGAAC CCAAGGCGTA	1140
GCCTCCTACC ATGGCAATAG TTGCTAAAAT AAGGCCTAAC CACTGACTTT TTCCTTTAAA	1200
TCCTGCGAAA AATCCCTGCA AGCCATGGTT GACCAAGCTA AAGAACATCC ACTGAGGGTA	1260
GCCTGATAAG AGGTCAATCA AGAACTTGC TAGTCTCCG ACTACCGCTC CTTACGACT	1320
ACCAAAGTAA AAGGCCGCAA AGAAGACACC AGCATCTAAA AGAGTTAGAA TTCCTGTAGG	1380
TGTTGGGATT TTTAAGAAAT AACCTAGAAC CACAGAAAGG GCGGTAAATA GGGATACAAG	1440
GGCGATTTTA GTTGTTTTGT TTTGCTTCAT ATTGTCTTAC TCCATACTGA TCTGCTGTG	1500
CAATAGCACG ATAAACGAAA GCCTTAGAGC TTTCTACTGC TGGCAAAAGT TTATCACCTT	1560
TAACCAGGTG ACTGGCAATG CTAGAGsCAA AGGTACAACs TGCACCAGCA TTTTGGCCTT	1620
GGATAACTGG ATTTTCTAGG ATAGTAAAGG TCTGTCCATC ATAAAAGACA TCCACAGCCT	1680
TGTCCTGACT AAGACGATTG CCTCCCTTGA TAATGACTGt GCGCTCCTA AATCATGCAA	1740
TTTCTGCGCT GCAGTTTCA TGTCTTCCAA GGTTTTAATT TCCTGACCGG ATAATAATTC	1800
TGCTTCTGGG AGATTAGGCG TAATCACACT GACATAAGGG AAAAAGCGAA TCAACTCTTG	1860
GCAGAGCTCA CTGACAGCTA CATCATGCGT TTCCTTGCAG ACCAAGACAG GATCCAACAC	1920
CACAGTACT CCTGGGCGTT GTTTGATAAA GTCCAAGGCC TTCTCAGCCA CGCTGACAGT	1980
AGGGAGAAGA CCAATCTTAA TTCCCCCAA TTCCACATCA CGCAAGCTAT CTAATTCATG	2040
TTGAAAAATG GTATCATCAG TTGGAAAGAC TTCAAATCCT TTTTCTGTCA AGGCTGTCAA	2100
ACAAGTCACT GCTACAAACC CATGCAAGCC GTTCAAGGTA TAGGTAGCCA AATCAGCTGA	2160
CAGTCCACCA CCACTAAAAA TATCATTTCC AGAAAGTGCT AAAATACGAT TATTCTTCAT	2220
AACGAATCTC CTTTAAATAC AAACCATTTG GTGCTGCAGT GGGACCTGCA AGTTGCCTGT	2280
CCTTCTTCTC CAAGATGAGA TCAATCTGCT CTAATGGCAT GCGGTGTGTA CCGATTTTGA	2340
GAAGAGTCCC CACCATATTG CGAATCTGTT TATACAAGAA ACCATTTCTT GAAAAGGTAA	2400
AGGTCAAAAA TTGTCCTGTC TCATCGACTA TTAACTAGC TTCTGTGATG GTGCGAACCT	2460
TATCTCTTAC ACTAGTCCCA GAGGCTGTAA AACCGGTAAA ATCATGGGTT CCCTCTAGCT	2520
TTTTGATTGC AATCTGCATT CGTTCACAT CGAGTGGGTA GGGAAAGTGG GTGGCATAGT	2580
GACGGCGCAT CGGATTTTGT GGACGTCCTC TATCCACAGT AAATCATAG GTCTTGCTAT	2640
GCTTGGCATA ACGGCAATGA AAATCATCTG CCACAAGCTC AATCGAAATC ACATCAATAT	2700

334

CTTCAGGAGA CTGGGTATCC AAGGCAAAAC GGAGTTTCTC CTCATCCATC TGATAAGGCA	2760
GGTCAAAATG AATCACCTGT CCCAGGGCAT GAACCCCACT ATCTGTCCCTA CCAGCACCGT	2820
GAACAGTAAT GGCTTGCCCT TTATTTAATC TGGTCAAGGT TTTTTCATTT TCTTCCTGAA	2880
CGCTACGCGC ATGAGGCTGG CGCTGAAAGC CAGCAAAGGC ATAACCATCA TAGGAAATAG	2940
TTGCTTTATA TCTCGTCATA GCCTCTATTT TATCAAGAAA TTAGTCTGTA AACAAGGACC	3000
TAAAACAAAT ATTGTATGGG TATAAAAATC TCATACTCTT CGAAAATCTC TTCAAACCAC	3060
GTCAGTTTCC ATCTGCAACC TCAACACACT ATTTTGAGCA ACCTGCGGCT AGCTTTCTAT	3120
ACTAGATTGA AATAAGATAT GAACAACCTCT ATTAGGAAAG TCAAATTAAT TTCTAGAAAT	3180
ATTTTAGCAG CTACAGCGTA CTATTCCAAA CTCAATCAAC TATAGTTTGC TCTTTGATTT	3240
TCATTGAGTA TCAAAGAAA AACTTAGGAA TCAATCCTAA GCTCTCTTCT GAAGTAGGTA	3300
CATGACAAAG ATAGAGATTA CAATCAACCA ACCTCCTAAG ATACTAAAGA CCAACATCCC	3360
ATTGTGAGTT AGTAAGCCAA TTGCACCTAG AACGAATGGG GTCGTAAAGG CTCGAAACT	3420
ACAGCCTAAT ACAGCAAATG AAGTTGCTTG ATTGAGGAGT TTAGCTGGAA TTCGTTTCTAGA	3480
GACAAGTTGA AAGACCGTCG TCAAGACTAC ACTATAGGCA AATCCAGCCA GAACACTTCC	3540
TGCTACTACC ACCCACAAGG ATGAAGACAA GGCAATCAGC ATTTGCCCA AGCCAAAGGT	3600
AATACCAGAC CAGAGGAGCA GTTTCTCTTT AAAGATAGAA ATCAAGAAAG AAAAACTCAC	3660
CCCAGCCACA ATCCCGATCA ACTGCATGAT ACTAAGAACA AAACCTAGATA ACTGGGCATC	3720
CCCCAATCCT CTTTCCACCA TCAAACCTGG AATACGGATG GTAATAGCTG TATTGGTACA	3780
AACTACAACCT GCCGCTTCGA TAGCTAAGGT AAAAATCAAG CCTTTCATTT CTCGAGTTAA	3840
ACGACTTGCT TCCTTCGCTC TTTTCTTGAC TTCTTTCTTT GATTTTCCAT AAGGGACAAA	3900
GAGCAGATAA AGGGGCAGCA CCAAAAATCC AGCACTATAG GCTAGAAAGA TAGCTGTCCA	3960
ACCAAAGGCC AACAACTGAC CGACGGCCAA GGTAATGAGA GAAGCTCCAA CGACCTCTGC	4020
AGAAGCGCGT AGCCCTAACA TCTGAATTCG CCTTTTCTCT TGGTAGCGTT CACTGATAAT	4080
AGAAATGGCC TTGGCATTGA TCATCCCAAG ACCCAAACCA AAGAGAAGCC GTGTTCCAAA	4140
GACAAAGGGA TAGGCTTGGT ACCAGAAGGG AGCTGTACCG CTCAATGATA AAATCAGCAA	4200
GCCCAAACTA ATCTGTAAGC GCTCAGGAAA TATTTTCTCT AAGAAACCAT TTAGCAGTAA	4260
CATCATCATG ATTCCAAAGG AAGGCAAGCT CACCAAGAGC TCAATTTGTT CCTTAGAATA	4320
ACCCTGATAA TAGTCAAACA TGGCTGGTAG GGCACCTCGAA ATGGAAAAGG AGGTAATCAA	4380
AACGAGGGAG AGAGCCAAAA TGCTGGCCCG TTCTAAAAAT TGTTTCATGA AATCTCTTTC	4440
TATATTTCTC TTAATCTTCT ACTTTTGTGA TAGTTATCAA ATAAGCAAGA AAAGAAGAAG	4500

335

CCTCATTGGT TTGTAGACTC CTTCTTAAAT TCGAAAATGA ATCCCTTGTA TCTTATACTC	4560
AATGAAAATC AAAGAGCAAA CTAGGAAGCT AGCCGCAGGT TGTTCAAAAC AGTGTTTGA	4620
GGTTGCAGAT GGAAACTGAC GTGGTTTGAA GAGATTTTCG AAGAGTATTA GGATGACTTT	4680
CTCTTGATTT GCTTGATAAA GTAGAAAATA AATCCTGCTA CCATATAGGC AACAAAGATA	4740
ATCAGACACC ACTTAAACAC AACATTCCTA CCCTTGTTCA CATTCAAAAA GAAGTAAGGG	4800
AAAGGATTAT CCTTGGCATT TGGAATATTG AGTTTTAGAA CCAAGCCATT AAAAAGAGCA	4860
AACATCATAT ACAGAAAGGG TAAAATGGTC CACACTGCTG GATCCCAAAT CTGTGATTGA	4920
CCCTGTTTGT CAAAAAGAG GGTATCCGCT AAAAACCAGA TGGGAACGAT ATAGTGGCAA	4980
AGGAAATTTT CTAGGGTATA GAAATTAGTC GCAATGGCG CCAAGAGGAA ATGGTAAATC	5040
ACACAGGTAA TCATGATACT CATGGTGACC CCACCTTTTA AGCGCAAGAG ACTTGGCCTT	5100
TGCCAATTTT CACCTACACG GCTCATAACC TTTAGAAGAT AAAGGGTAAA AATAGTTACC	5160
AAGAGTTGG ACAGAACCGT GTAATAGAGA AGCATCCCA AACCACCATG CTTAGTAAAT	5220
TCAAGATAAA CTCCCGTAAA AGCCGCTAGA AACAAGAAGA TACGGCTATA AAATACAAGT	5280
TTTAGTGT TTGACATGCT TAAATCTTCC TCACAACTC TGATTTAAGT TTCATGGCAC	5340
CAAAACCATC AATCTTACAG TCGATATTGT GGTCGCCTTC TACGATGCGG ATATTTTCA	5400
CGCGCGTCCC TTGTTTCAAA TCTTTTGGCG CACCTTTTAC TTTCAAGTCC TTGATGAGAG	5460
TTACTGTATC ACCATCAGCC AATTTATTTC CGTTGGCATC GATAGCGACA AGACCTTCTT	5520
CTACTTCTGC AACTTCAGCA GGATTCCACT CATGAGCACA CTCTGGGCAA ACCAGTAGGG	5580
CACCGTCTTC GTAGACATAC TCTGAGTTAC ATTTTGGACA ATTTGGTAAA TTGTTTCATGG	5640
TTTCTCCTTA TCATCATTC AATTCTTTG AAAATCAAAA TTTCTCGAAC AGCAACTATT	5700
ATACCTAAA ATCAGCATTT TGACAAATT AGAAAAAAC CGATATCAAT CTATCGGCTT	5760
TTCTACATTT ACATTCCTTT TTCAGCTTCT GCTTTGATTT TTTCAACTAC TTCTTGAATG	5820
TTCAAACCAG TTGTATCAAG GTAGACAGCA TCCTCTGCTT GTTTGAGAGG AGAAGTCTCA	5880
CGATGACTAT CCTGTAGTC ACGCGCAGCA ATTTCTTTT TTAGGGTTT AAGGTCTGTT	5940
TCAATTCCTT TGGCAATATT TTCCTTGTA CGACGCTCTG CTCTCTCATC AACAGAAGCT	6000
ACTAGGAAAA TTTTCAATTC TGCTTGTTGC AATACAACAG TTCCAATATC GCGACCATCC	6060
ATGACAATCC CGCCTTGCTG GGCAATTTCT TGTTGGAGAG AAACCAGTTT CTCACGCACT	6120
TGAGGAATTG CTGCAATAGC AGAAACATGA TTGGTCACTT CATTTTCACG GATAGGATGG	6180
GTAATATCCA CATCTCTAC AAAACAAGC TGGTCTCCAG TTTCTGAACG TCCAAAGCTG	6240

336

ATTGGATGCT GGTCCAACAA GGCTAGAAGG GCTTCGACTT CTTCAACTCC TAATTGGTTC	6300
TTAAGAGCCA TATAGGTCGC TGCACGATAC ATAGCTCCTG TATCAAGGTA GGTGAATCCA	6360
AAATCCTTAG CAATAATCTT TGCACCGTA CTCTTACCGC TGGAAGCAGG ACCATCAATA	6420
GCAATTTGAA TTGTTTTCAT ATCGGCTCCT ATTTTATTTT TATAACATCA CCTGGATTAG	6480
CAAACCAAGA TCCTGTAGCC ATGTGCCCAG GATTC AAGGC CTCTAACTGA GCAATGGAGA	6540
TTCTGTCACG AGCGGCAATA GCTGCTTCCC CTCTCCTGC GAGAACTTTA ATCGTTCCTT	6600
CAGGATTAGC AGCTTCTTCT GAACTACTAG AAGTAGATTG TGGCTCTGAA CTCTGCTCAG	6660
GCTGAGAACT ACTTGAAGAT GAGATTGTGA CTACACTGGC ATCAGAATCA TGAAAGCCTT	6720
TTAAGGCTGC TGTGCGATTA CTCCCCCCG ATGATAGATA GATGAGAACG ATGACCATCA	6780
CCACCACAAT TACAAAGAAA ATACTAGCTA GGATCGTCAA AATACGATTA GCCATCCTAT	6840
CAGCCCTCC GTGGTTTCGA TGCCGACGCT CTGCTCTTGA TTCTTCTTGA TCATAGATAT	6900
CTTCTTGCCA CGGTCTTCTT GCCATACCTT ACTCCTTGT TTTTCTTACT TTTCTTATTA	6960
CAATATAAAT ATGAACATGA AAATCACACT TATACCTGAA CGATGTATCG CCTGTGGGCT	7020
TTGCCAAACT TATCTGATT TATTTGATTA CCACGATAAT GGAATCGTGC GTTTTTACGA	7080
TGACCTGAC CAACTGGAAA AAGAAATTC TCCTAGTCAG GATATCTTAG AGGCTGTAA	7140
AAATTGCCCA ACTCGCGCCC TGATTGGAAA CCAGGAAGCC TAAATCAATG GCGATAATCC	7200
ACTCCCTCTA GTTTAGCACA TTTCCATGTA AAATTATAGT CTTTTCACCT TATTTTPTTC	7260
TGTAATAATCA GGAAGGTCAC TTTTPTCTTT GATAAGATAA AGTGGTCTTT TTTTAGTCTC	7320
TAAATAAATC TTACTGATAT ACTTGCCGAG AATCCCAATG GTCAAGAGTT GAATGCCTCC	7380
AAGAAAGAGA ATAACAGCCA TCAGAGAGGT CCAACCAGAT GTCGGATTGC CCAAAATGAG	7440
GGTCCGAACC ACAACAAAA AGGTCATCAG CAGAGAAAGA AAACAAGATA GGAGACCAGC	7500
TACAAAGGCT ATAATCAAGG GAAAACTGA AAAATTAATA ATCCCTTCAA TGGAGTAGAA	7560
AAAGAGTTC CTAAACTCC AACTTGCTT GCCAGCCTGC CTTTCGACAT TTGGATAGTC	7620
CAAATAGTAG GTTTTGAAAC CCACCCAGGC GAAGAGCCCC TTTGAAAAAC GATTGGACTC	7680
GGTCAAGCTT AAAATGGCAT CGACTACAGA CCTTCTCATC ATACGAAAAT CACGGACACC	7740
CGACGGCAGA GCTACTGGGC TGATTTTTTG CATGAGGCGA TAAAAGAGAA CAGCACAGAA	7800
ACTGCGAAAG AAGGGTTCTC CCTCCGACT AGTTCTCCGT GTCCCAACGC AGTCCAAGTC	7860
TACATTTTTG TCTAATACAT TTTTCATCTC AAACAACATA CTAGGAGGAT CTTGGAGGTC	7920
TGCATCCATC ACCACCACCA AATCTCCTGT CGCATATTGC AAGCCTGCAT AAAGGGCTGC	7980
TTCTTTGCCA AAATTTTCGAG AGAAAGAAAT ATAATGGACT GCCGATT TTT GCTCCCGATA	8040

337

GGCCTTTAAG AGTTCCAAGG TCCCATCACT TGATCCATCA TCGACAAAGA CATACTCGAT	8100
TTCTGTTTCC AAATCTGGAA GTAAAGCTTC CAGAGCCTGA TAAAAAGAG GAAGTACTTC	8160
CTCTTCGTTT AAACAAGGGA CGATGATTGA AATCATCATC TTAGTCTTCA AATCCATTTG	8220
GATGCTTGCT TTGCCAACGC CATGCGTCTT CACACATTTG GGTGATGTCG AGTTCTGCTT	8280
CCCAACCGAG TTCGTCTTA GCTTTTGCCG GGTCTGAGTA GCAGGCAGCG ATATCACCTG	8340
GGCGACGTTT TACGATGCGG TAAGGAATAG GACGGCCAC CGCTTTTCC ATGTTTGGGA	8400
TAATTTCAAG AACTGAGTAA CCTTTACCAG TTCCAAGGTT ATAAACGTTT AGTCCTGAAC	8460
CTTTTGGAT TTTTTCAAA GCTGCAACGT GACCCCTAGC CAAATCGACA ACGTGGATAT	8520
AGTCACGAAC ACCTGTCCA TCTTCCGTAT CGTAATCGTC TCCAAACACT TGCATTGCT	8580
CTAATTTTCC AACGGCTACT TGAGTCACAT ATGGCAAGAG ATTGTTTGGGA ATACCGTTTG	8640
GATTTTCTCC CAAATCACCA CTCTCATGGG CTCCGATTGG GTTAAAGTAA CGAAGCAAGA	8700
CAACATTCCA TTCTGAGTCT GCTTTGTAAA TATCAGTCAA AATTPCCTCT AGCATGAGCT	8760
TAGTACGACC GTATGGGTTG GTCACTGAAA GTGGGAAATC TTCCAAGATG GGCACGTGT	8820
GCGGATCCCC GTAAACTGTC GCAGAAGAAC TGAAGATGAT GTTTTACAG TTGTTTCTT	8880
CCATGGCTTT CAAAAGGCTG ACAGTTCCAG CGATATTGTT GTCATAGTAG GCAAGAGGGA	8940
TACGTGTTGA TTCGCCAACA GCCTTCAAAC CAGCAAAGTG AATGACACCA GTCGGTCTT	9000
CCTGCTTGAA AATATCTCTG AGGGTATCTG TGTCACGAAT ATCTGCCTCA TAGAAAGGAA	9060
TCTCAACTCC TGTGATTCCT TCAACAACTT CTAAACTCTT ACGATTGCTA TTGACAAGAT	9120
TATCCACCAC AACAACTGA TGACCTGCTT GGATCAATTC AATAACAGTG TGGTTTCAA	9180
TAAAACCGGC ACCACCAGTT ACCAAAATCT TTTCTTGCAT CTTTTTTCCT CGATTCTCAG	9240
ATTATTTTTT CTTATTTTAC CATTTTGAC AGGGAATGTC ATTTGCCATC CTAAACTACC	9300
TGATAAAATT TCAGTAAAT GCTTATACTC TTCGAAAATC CAATTCAAAC TACGTCAACG	9360
TCGCCTTGCC ATGGGTATGG TTA CTGACTT CGTCAGTTCT ATCCACAACC TCAAAACAGT	9420
GTTTGTAGCT GACTTCGTCA GTTCTATCCA CAACCTCAA GCAGTGCTTT GAGTAACCCG	9480
CGGCTAGTTT CCTAGTTTGT TCTTTGATTT TTATTGAGTA TTATTCGCTT TTTACTCGTT	9540
TGACATAGTT TTCAATTGGG TAATTTAGAG GGTCCAAGGT CAACTCCTTG TCTTGGATCA	9600
GTTGGGCTAG ATGGTAACCA ATGATAGGAC CAGTTGTGAG GCCTGATGAA CCTAGTCCAC	9660
TGGCTGCATA GACACCAGTT AAGTCAGGCA CCTGCCCAA GAAAGGAGAG AAATCACTGG	9720
TGTAGGCACG GATTCCAACA CGCTCAGATT TTGAAGTAGC TTCAGCCAAA ATCAGATAGT	9780

338

GAGTCAAGGT	GGCCTCCTCC	ATTGTGTGGA	GCAAGGTTTC	ATCTACCGTC	AAATCAAATC	9840
CCATGTCATT	TTCGTGGGTA	GCGCCTAAGG	ATAATTTCCC	ACCTGCAAAG	GGAATCAAAT	9900
CCCACTCCCC	TTCTGGCATG	ACAACAGGGT	AATCTTCCAT	GTCTTGGGCA	AGCTGATAAT	9960
CTCGTAGTTG	TCCTTTTGA	GGACGGACAT	CCACTTCATA	ACCTAAAGGC	TCTAACATGT	10020
CCCCCAACCA	AGCTCCCGTC	GCCAAAATAA	CCTGCTCAAA	CTCCTCTTCA	CCAATCTGGT	10080
AGCCTGATGC	TAACGGTGTC	AGAGTCACTT	TTTCTTTGAC	CAGCTTGACA	TGACTGACTT	10140
CCAGCAAACG	AGTCACTAAA	AGTTGGCCAT	CTACTCTCGC	TCCACCAGAA	GCATAGAGCA	10200
GGCGGTCAAA	TCCCTGCAAA	CCAGGGAATA	ATTCATTAGC	TGAGGCTTGG	TTCAGAATGG	10260
CTAATTGCCC	TATCAAGGGA	GATTCTTCTC	TGCGCTGGAG	GGCCAGTTGA	TAAAGTTCTT	10320
CCAAATTGGA	TTCATCCTTT	TTCAAGAGAA	AGACTCCCGA	ACGCTGGTAA	AAGTCGATTT	10380
CTTGTCCTGA	TTTCTCTAAA	TCAGCTAATA	AATCCACATA	AAAATCAGCC	CCCAAGCGCG	10440
CCATCTTGTA	CCAGGCTTTA	TTACGGCGTT	TGGAAAACCA	AGGACTGATA	ATTCTTGCTG	10500
CGGCCTTGGT	GGCTTGACCT	TGCTCATGCT	CAAAAACGGT	CACCTCTAGG	TCACTTTCTC	10560
TCGAGAGGTA	GTAGGCAGCT	GTGCTCCCA	CAATTCTCTG	TCCAATAATG	GCAACTTTTT	10620
TCATTGTCTT	CACTTTCTAA	CTAGATATGA	TGGAAAGGAT	TGGTTGATGC	CTGACTAGGC	10680
AAGATATCAA	TAGACCACCC	CTTATCTTCC	TTCCATTGAC	TAAGAAGTGC	TGCGATTTTT	10740
TCTACAAAAA	TCACTTCGAT	ATAGTGACCT	GGGTCCAATG	CAAGCAACCC	ATCAGATAGC	10800
ATATCCTGAG	CAGTATGGTA	GTAGATATCA	CCAGTGATAT	AGACATCTGC	CCCCTTTGCC	10860
AAAGCATCCT	TATAGAAAGA	CTGCCCGCTT	CCACCACAAA	TTGCTACTCT	TGAAATAGGC	10920
TTCTGCAAAT	CATCCTCTTG	ATAATGCACC	ATTCGAAGGC	TATCTAGGTC	AAAGACTTGC	10980
TTGACCTGTT	GGGCCAATTC	CCAAAATGTC	TGAGGCTGAA	TATTCCCAAT	ACGTCCAATT	11040
CCACGTCTTG	GACCTGTTTC	CTGCAGATAA	GTGCTCTCCT	CGATTCTTAG	CATCTGACAA	11100
AACCAGTCAT	TGAGCCCAT	TTCAACGATA	TCAATATTGG	TATGGCTGAC	ATAAACTGCG	11160
ATATCATGCT	TAATCAGGTC	GATGTAAATC	TGATTTTGCG	GACGGCTGGC	AAGCAAGTCC	11220
TTGATAGGAC	GAAAGATAGG	CGCGTGCTTG	ACGATAATCA	AGTCCACACC	CTTTTCAATG	11280
GCCTCTGCCA	CTGTCTCTTC	ACGAATATCG	AGGGCAACCA	TGACCCTTTG	GATACCCTTG	11340
TCTAAAGTGC	CAATTTGCAG	ACCACGGCTG	TCTCCCTCCA	TAGAAAATTC	CTGAGGGCAA	11400
AAGGCTTCAT	AAGCTTGGAT	CACCTTCACT	GCTAACATGG	AGCACCTCCT	TGATAGCTTG	11460
AATCTTATCT	ACTAGAACTT	GACGTTCTTC	CAGATTTTTT	TCTGGGATTT	GTCCGAGGGC	11520
GAACTCTAGC	TTCTCAGCTT	CTTTTGGCCA	TTTTTGGACA	AATACTGGAC	TGACTTCTTT	11580

339

GGACAAGAAG GGACCAAAGC GAACATCACT GGCTGATAGC TTCATTTGTC CTGCTTCCAC	11640
CACCAAAATC TCATAAACT TTCCAGCTTC TTCTAAGATG CTTTCTGCTA CAATCTGGAA	11700
TCCATGATCC TGTAGCCAGA TACGCAAGTC GTCTTCACGA TTATTTGGGCT GGAGGATCAA	11760
ACGCTCTACA TTAGCTAACT TCCCCAAACC TTCTTCTAAA ATCCTAGCAA TCAAACGACC	11820
ACCCATGCCA GCAATGGTAA TGACAGACAC TTGGTCAGTC TCTTCAAAAG CTGCCAAGCC	11880
ATTGGCTAAA CGGACTTGGA TTTTCTCCTT TAGGCCGTGA GCCTCAACAT TTTTAACCGC	11940
AGACTGATAG GGACCTTCCA CCACCTCACC TGCAATAGCG CTTTGTGATT GGCTCTCTC	12000
AACCAACTCG ATAGGCAGAT AAGCATGGTC ACTTCCCACA TCTAGTAAA TAGCCCCCTG	12060
TGACACAAAG GAAGCTACCA ATTCTAATCT CTTTGAAATC ATCTTCTCTC ACTTTCCAAA	12120
ACTCTATTAC CTCTTATTAT ACCACATTTC AATCTTCAAC TTCCCAGTAA TATAAGCACC	12180
TCTGGCGAAA GAAGTTTCAA TGTCTTAAAG TAATAAGTGA ATCCAATTGA AAGATTTTAA	12240
ACAATTTGCA AAAATGTCAA AAAATAAAAA ATAAACAGTT TATTCAGAAA ATTCTTGACA	12300
TATAAAAACA CATGGTAGAA TATAATTAGA AAGTTAGAAA AAATAAAAGT TTGACTAAAA	12360
TTTGTATTTG AAGGTGGTGT TCAGATAAGA AATTTAGTCA GACGAACCAC GAATTTGCTC	12420
TATGCTTTCT GGAATTATC ATAACAGGAG GATACAGTCA TGGAACAAAC ATTGTTTGAA	12480
TTAGAACTAC TTCCAGAGGA AGATATCATT GTCACAGGTC TCCCTAAGTA TTGTTCTTTT	12540
ACTTGTTTAA TTACAGGTCG CTAGTTATAT TTTATATAAA ATAAGTAGCT TTACTTACGG	12600
AATAGGCTAG TGCTGTGTCT CTAGCCTATT TTAATAATTA GGAGTTTGTT ATGGATTTAT	12660
TAGAGAAAGA ATGTTTAAAA TGTGATAAAA ATTTCCAACA GGGTGATATT TGGAATTACT	12720
ATTATTTATC AGATAAGATG CCTGCACAAG GGTGGAAAAT ACACATAAGC TCCCAAATAA	12780
AAGACGCTGT AAATATTTTT AAGATTGTGT ATAAACTATC CCAACTAAAT AATTGTAGCT	12840
TTAAAGTTGT TAAAAATTTA GAGGAATTAA AAAAAATTAA TTCCCCTAGG GAAATGAGCC	12900
CTACTGCTAA CAAATTTATA ACTCTATATC CTAAGTCAGA ATCTGAAGCT AAGAGTATGA	12960
TTTGTAATCT TACGAATAGA CTGTCAGAAT TTAAGGCTCC AAAAATACTA TCTGACTATC	13020
AATGTGGAAT GCATTCTCCA GTTCATTATA GATATGGGGC TTTTTTAAAA AAACAAGCTT	13080
ATGATGAAAA AAATAAAAAA GTCATCTATT TATTGCTAGA TGAAAAAAGG AAGAACTATG	13140
TAGAAGATAA GAGACAAAAT TTCCCTAGTC TTCCTAGCTG GAAAATGGAT TTATTTTCAG	13200
AAGAAG	13206

(2) INFORMATION FOR SEQ ID NO: 34:

340

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13104 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

```
CCGGATCCAG CGAAAAATAT GCTCTTTGAT GCTGTAAGTG GTCAAAAAGA TGCTAAAACA      60
GCTGCTAACG ATGCTGTAAC ATTGATCAAA GAAACAATCA AACAAAAATT TGGTGAATAA      120
AAAATTTGTT CAAGGGGGGT GGAAATCAAA TCCCCCTTTG AATTTATCAA TAGAGACACA      180
AATAATTTAG CTTTCTTATA AAAAAGTAGT ATCCTATGAA AGGAGTTAAT ATGGAAAAGC      240
AACAACTAG TAAAGCAGCC CTGCTGTCTA TCATTCTGG GTTAGGACAG ATTTACAATA      300
AACAAAAAGC CAAAGGTTTT ATCTTCCTTG GTGTAACCAT CGTATTTGTC CTTTACTTCC      360
TAGCACTTGC AACCCCTGAA TTGAGCAACC TCATCACTCT TGGTGACAAA CCAGGTCGTG      420
ATAATCCCT CTTTATGCTG ATTCGTGGTG CCTTCCATCT AATCTTTGTA ATCGTTTATG      480
TACTCTTTTA TTTCTCAAAT ATCAAAGATG CACATACGAT TGCAAAACGC ATTAACAATG      540
GAATTCAGT TCCACGCACA CTCAAAGACA TGATCAAAGG GATTTATGAA AATGGCTTCC      600
CTTACCTCTT GATCATTCCA TCTTATGTTG CCATGACCTT CGCGATTATC TTCCCAGTTA      660
TCGTAACCTT GATGATCGCC TTTACCAACT ACGACTTCCA ACACCTGCCA CCAAACAAGT      720
TGTGGACTG GGTGGGTTG ACCAACTTTA CAAACATTTG GAGCTTGAGT ACCTTCGGT      780
TGCCCTTGG TTCTGTCTT TCTTGGACTA TCATTTGGGC TTTGGCAGCT TCTACTTTAC      840
AAATCGTAAT TGGTATCTT ACAGCTATCA TTGCCAACCA ACCATTTATC AAAGGAAAAC      900
GTATCTTTGG TGTATTTTC CTTCTTCCTT GGGCTGTCCC AGCCTTCATC ACTATCTTGA      960
CATCTCAAA CATGTTTAAC GATAGTGTG GTGCTATCAA CACTCAAGTA TTGCCAATCT      1020
TGGCTAAAT CCTTCCTTC CTTGATGGAG CTCTTATTC TTGGAACA GACCCAACCT      1080
GGACTAAGAT TGCCTTGATT ATGATGCAAG GTTGGCTCGG ATCCCATAC ATCTACGTT      1140
TGACCTTGGG TATCTTGCAA TCTATTCCTA ACGACCTTTA CGAAGCAGCT TATATTGACG      1200
GTGCCAACGC TTGGCAAAAA TTCCGCAACA TCACTTTCCC AATGATTTTG GCTGTTGCGG      1260
CACCTACTTT GATTAGCCAA TACACCTTCA ACTTTAACA CTTCTCTATC ATGTACCTCT      1320
TCAATGGTGG AGGACCTGGT AGTGTGCGAG GTGGAGCTGG TTCAACCGAT ATCTTGATCT      1380
CATGGATCTA CCGTTTGACA ACAGGTACAT CTCCTCAATA CTCATGGCG GCAGCTGTTA      1440
CCTTGATTAT CTCTATCATT GTCATCTCAA TCTCTATGAT CGCATTCAG AACTACACG      1500
```

341

CATTTGATAT GGAGGACGTC TAAGATGAAT AACTCAATTA AACTCAAACG TAGACTGACT 1560
CAAAGCCTTA CTTACCTTTA CCTGATTGGT CTATCAATTG TAATTATCTA TCCACTGTTG 1620
ATTACCATTA TGTCAGCCTT TAAAGCAGGT AACGTCTCAG CCTTTAACT AGATACTAAT 1680
ATCGACCTCA ATTTTGATAA CTTTAAAGGC CTCTTCACTG AAACCTTGTA CGGTACTTGG 1740
TACCTCAACA CTTTGATTAT CGCCTTAATT ACCATGGCTG TTCAAACAAG TATCATCGTA 1800
CTTGCTGGTT ATGCTTACAG CCGTTACAAC TTCTTGGCTC GTAAACAAAG TTGGTCTTC 1860
TTCTTGATCA TCCAAATGGT GCCAACTATG GCCGCTTTGA CAGCCTTCTT CGTTATGGCG 1920
CTTATGTTGA ACGCCCTTAA CCACAACTGG TTCCCTCATCT TCCTCTACGT TGGTGGTGGT 1980
ATCCCGATGA ATGCTTGGCT CATGAAAGGC TACTTCGATA CAGTGCCAAT GTCTTTAGAC 2040
GAATCTGCAA AACTAGACGG TGCAGGACAC TTCCGCGCT TCTGGCAAAT TGTTCTACCA 2100
CTTGTTCCGC CAATGGTTGC CGTACAAGCT CTCTGGGCCT TCATGGGACC TTTCGGGGAC 2160
TACATCCTCT CTAGTTTCTT GCTTCGTGAG AAAGAATACT TTAAGTTGCG CGTAGGTCTC 2220
CAAACCTTCG TTAACAATGC GAAAACTTG AAGATTGCCT ACTTCTCAGC AGGTGCTATC 2280
CTCATCGCCC TTCCAATCTG TATTCTCTTC TTCTTCCTAC AAAAGAACTT TGTTTCAGGA 2340
CTTACAAGTG GTGGCGACAA GGGATAATTT ATCCCCGCCA CCCTTTTCA TTTTATACTC 2400
TTCGAAAAATC TCTTCAAACC ACGTCAGCTT TATCTCCAAC CTCAAAGTTG TGCTTTGAGC 2460
AACCTGTGGC TAGTTTGAC TTTGATTTTC ATGATTATT AGCAATTGTC ACTGTAAATA 2520
ATATCCTTGT AGCAAGCAAT TTTTCTCCTA GACTTGAAAT AAAGCGCATT TCTCTATATA 2580
ATAATACTCA TATAGAAAAC ACCTTTTAGA AAGATACCTA TGCTTCCATA TCCATTTTCC 2640
TATTTTCAA GTATTTGGGG GGTTCGTAAG CCCCTGTCCA AACGTTTCGA GCTCAACTGG 2700
TTTCAACTTC TCTTTACCAG TATCTCCTT ATCAGCTTGT CTATGGTACC CATTGCTATC 2760
CAAAACAGCT CCCAGGAGAC CTATCCGCTA GAAACTTTTA TCGATAATGT CTATGAACCT 2820
CTGACAGATA AGGTGTGCTA GGATCTCTCT GAACATGCTA CAATTGTCGA TGGCACATTA 2880
ACTTATACTG GAACAGCTAG TCAAGCCCTT TCTGTTGTGA TTGGTCCAAG TCAAATCAAG 2940
GAATTACCTA AGGACTTGCA ACTGCATTTT GATACAAATG AGCTAGTCAT CAGCAAGGAA 3000
AGCAAGGAAC TGACCCGCAT CTCTTACCGA GCCATTCAGA CTGAGAGTTT CAAAAGCAAA 3060
GACAGCTTGA CCAAGCAAT TTCTAAAGAC TGGTACCAAC AAAATCGTGT CTATATCAGC 3120
CTCTTCTAG TTCTCGGTGC GAGCTTCTC TTTGGTTTGA ATTCTTTAT CGTCTCTCTT 3180
GGAGCTAGCT TTCTCCTTA TATCACCAA AGATCAGGCC TCTTTTCATT TAATACCTTT 3240

342

AAAGAGTGCT ACCATTTTAT CTTGAACTGT TTAGGATTGC CGACTCTGAT TACACTTATT	3300
TTGGGATTAT TTGGCCAAAA TATGACAACC CTGATTACTG TACAAAATAT TCTTTTGTGTT	3360
CTGTATCTGG TCACTATCTT TTATAAAACA CATTTCCGTG ATCCAAATTA CCATAAATAG	3420
GAGATTTTTA TGCCCGTTAC GATTAAAGAC GTGGCCAAGG CTGCTGGTGT TTCGCCTTCA	3480
ACCGTAACCC GTGTATTCA AAATAAATCA ACCATTAGCG ACGAAACAAA AAAACGTGTT	3540
CGCAAAGCTA TGAAGGAACT CAACTACCAC CCAAACCTCA ACGCTCGTAG CTTGGTAAGC	3600
AGCTATACTC AGGTATATCGG ATTAGTTCTT CCTGATGACT CAGACGCCTT CTACCAGAAT	3660
CCTTTCTTTC CATCGGTTCT ACGTGGCATC TCTCAAGTCG CATCTGAAAA CCACTATGCC	3720
ATTCAGATAG CAACAGGGAA AGATGAGAAG GAGCGTCTCA ACGCTATTTT ACAAATGGTC	3780
TACGGCAAGC GTGTAGATGG GCTAATTTTT CTCTATGCCC AAGAAGAAGA CCCTCTCGTA	3840
AAACTCGTCG CAGAAGAACA GTTCCCCTTC CTTATCTTAG GTAAATCTCT ATCTCCTTTC	3900
ATCCCCTTG TCGACAACGA CAATGTTCAA GCTGGTTTTG ATGCGACTGA ATATTTTCATC	3960
AAAAAAGGCT GCAAACGCAT TGCCTTTATC GGAGGAAGTA AAAAGCTCTT CGTGACCAAA	4020
GACCGTTTAA CAGGCTATGA ACAGGCGCTT AAACATTACA AACTTACCAC TGACAACAAT	4080
CGCATCTACT TTGCCGACGA GTTCTGGAA GAAAAGGGCT ATAAATTTAG CAAGCGATTA	4140
TTCAAGCACG ATCCACAAAT TGATGCTATC ATCACAACCG ATAGCCTCCT AGCTGAAGGT	4200
GTTTGTAACT ATATTGCCAA ACACCAGCTG GATGTCCCTG TTCTCAGCTT TGACTCGGTT	4260
AATCCCAAGC TCAACTTGGC AGCCTATGTC GATATCAATA GTTTAGAGCT TGGTCGTGTT	4320
TCCCTTGAAA CTATTCTCCA GATTATTAAT GATAATAAAA ACAATAAACA AATTGTGTTAC	4380
CGTCAATTGA TCGCCACAA AATTATCGAA AAATAAGAGA CTGGGCAAAA AGTCGTTAAA	4440
AGCAAAAACG CATACTATCA GGTATTGAAA AAACCTGATA CTATGCGTTT TATTGTGGGA	4500
AGATTTACTT CCTTTTCTAC TGAAATTGAG TCTTTTCCCA AGATCTTTTT ATACTCAATG	4560
AAAATCAAAG TGCAAACTAG GAAGCTAGCC GCAGGTTGCT CAAAACACTG TTTTGAGGTT	4620
GTAGATGAAA CTGACGAAGT CAGTAACCAT ACCTACGGCA AGGTGAAGCT GACGTGGTTT	4680
GAAGAGATTT TCGAAGAGTA TTAATCACTA ATTATCTATC TCAACAAATC TTCCTAGAAT	4740
ATGAACATTT TCCGAGACAG AGACAAAGGA GCTTGGATCC ACTTGTGTCA TAATCTGTTT	4800
AAATTCATTA AACTCTGCAC GTGTAATGAC AGTGATTAAA ACTGCCTTTC TCTCGTGATT	4860
ATAGGTTTCT TCTGCATCGT GGATCATGGT TGCTCCGCGG TGCAATTTTT TATGGATTTT	4920
TTCAATTACC TTCTCTGGAT GATTTGTAC AATCATGGCC TGCATACGCT TTTGCTTAGT	4980
AAAGACTGCG TCTGTCACAC GGCTAGAGAC AAAGATGGTA ATCCTAGAAT AAAGAGCGTA	5040

343

TTTCCAACCA AAGGTCAAAC CTGCTATCAG CATGATAGTT CCATTTACCA AGAAAGAAAT	5100
ACTACCGACA TTCTTACCCG TTTTCTTACG AATAGTCAGG CTGACGATAT CCGTCCCACC	5160
ACTGGAGATA TTGTTTCGAA GAGCAAAACC AATCCCCAAA CCCATAACAA CACCCCCAAA	5220
AAGGGAATTG ATAATGGGAT CCTCTGTCAA GGTGCCCACA GGGACAAACT GGATAAAGAA	5280
GGAACTCATA GATACCGTGA TAAAGGTAAA GACGGTGAAC TTATGGCCAA TCTGATACCA	5340
AGCTAAGACC ATCAAAGGGA AGTTAATGGC GTAGAAGCTT AGCGAAATCG GAATATGAAA	5400
ACCAAACCAG TGATTACTCA AGGCAGAGAT AATCTGTGCC AGACCTGTTG CACCACTCGA	5460
ATACACATGC CCTGGTTGGA AAAAGAAATT AACTGCTACT GCTGATAAAA AACCATAGAC	5520
CAGAGAGGCC GAAATCTTCT CATCATACTT TTCTCGAGAG ATACTTTGTA AGACACGTAA	5580
AATTTTATC TGATAAGCAA AGCGGCGCAG ATAATAGCGC CACCGCTTAA TTCGTTTGT	5640
TTGTTTCATC TTCTTCTACT TGTAAGCTGA GTTCCTCTAG TTGTTTGAGA GCGACTGTTG	5700
ATGGAGCTTG TGTCATTGGG TCAGTTGCCT TGTGTCTCTT AGGAAAGGCA ATGACTTCAC	5760
GGATATTTTC TTCTCCAGCA AGCAACATGA CAAAACGGTC AAGCCCGATA GCCAAACCAC	5820
CGTGTGGTGG GAAACCATAG TCCATGGCTT CAAGAAGGAA ACCAAACTGG TCATTGGCTT	5880
CTTCAGTTGA GAAACCAAGA GCCTTGAACA TGCGTTCTTG AAGGTCTTTT TGGTTGATAC	5940
GAAGGCTACC ACCACCAAGC TCATAACCGT TCAAGACGAT ATCGTAAGCA ATGGCACGAA	6000
CCTTAGCCAA ATCACCTTCT AATTCATGAG CAGTCTCTTC CTGTGGAAGT GTGAAAGGAT	6060
GGTGGGCGCT CATGTAGCGG CCTTCTTCTT CAGACCATTC AAACATCGGC CAGTCAACCA	6120
CCCAAAGGAA GTTGAACCTA TCATTATCAA TCAAGCCAAG CTCTTTAGCA ATACGTCCAC	6180
GAAGGGCACC CAGTGTGCA TTAGCCACTT CAAGCGTATC CGCCACAAAG AGAACCAAGT	6240
CCTTATCTTC AAGAACAAGC GCTGTTGTCA ATTCTTCTTG GATACCAGTC AAGAACTTGG	6300
CAACTGGTCC GTTTAATTCT CCATCAACCA CCTTGACCCA AGCAAGACCT TTGGCACCAT	6360
ACTGTTTGGC TACTTCCGTC ATCTTGTGCA TGTCTTTACG TGAATAGTTG TCCGCAGCTC	6420
CTGTGACCAC AATCGCTTTT ACAGCAGGTG CTTCTGAAAA GACTTTAAAG TCTACACCTC	6480
GGACCACTTC TGTCAAGTCC TGAAGCAACA TGTCAAAACG AGTATCTGGC TTGTCAGAAC	6540
CGTAAAGAGC CATAGCATCA TCGTATTTCA TACGAGGGAA TGGTAGCGTT ACTTCGATGC	6600
CTTTTGTTTC CTTTCATCAG CGCGCGATCA AGCTTTCTGT AATATCTTGG ATTTCTTGCT	6660
CAGTAAGGAA GGACGTTTCC AAGTCGACCT GAGTAAATTC AGGCTGGCGG TCTCCACGCA	6720
AGTCCTCGTC ACGGAAACAT TTAACGATTT GGTAGTAACG GTCAAAACCA GCATTTCATCA	6780

344

AGAGCTGTTT CGTGATTTGT GGACTTTGAG GAAGAGCGTA AAAATGCCCC TTATTAACAC	6840
GAGACGGCAC TAAATAATCA CGCGCCCTT CAGGCGTTGA CTTAGAAAGG AATGGTGTCT	6900
CCACGTCGAT AAACCTCCAAC TCATCCAAGT AGTTGCGGAT AGAGTGGGTC ACCTTGGCAC	6960
GAAGTTTAAG ATTTTCCAAC ATTTCTGGAC GACGAAGGTC AAGGTAACGG TAACGCAAAC	7020
GTGTATCGTC ATTTGCCTCA ATGCCATCCT TAATCTCAAA TGGTGTGTGC TTAGCTGTGT	7080
TAAGCACAAT AAGAGCTGTC ACGTTTAACT CAACCGCACC AGTTGGCAAC TTATCATTTG	7140
CTTGTCACGC GCAGCGACCT GACCAGTCAC CTCAATAACA AATTCGCTAC GAAGGcTTTC	7200
AGCTGTTGCC ATAACCTCTG CAGATACTTT TTCAGGGTTG ATAACCAACT GCATGATTCC	7260
TTCACGGTCA CGAAGATCGA TAAAGATCAA ACCACCAAGG TCACGACGAC GGCCAACCCA	7320
TCCTTTCAAG GTTATTTCTT GTCCGATGTG TTCCTCACGA ACACGACCAG CATACTACT	7380
ACGTTTCATT ATTTCTCTCC TCTTTTATTC TGTACTATT TTACCATAAA AGCGCAGCTC	7440
TTCATGAAAA TCATCAGAAA AGTTTGCCAG TCTTTAAAAG TCAGGTGAAA GCCCTAAAAA	7500
TTAGCGCTAA TACTCTCGA AAATCTCTC AAACCAGTC AGCGTCGCCT TACCGTATGT	7560
ATGGTTACTG ACTTCGTCAG TTTCTCTAC AACCTCAAAA CCATGTTTGT AGCTGACTTC	7620
GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA CCTGCGGCTA GCTTCCTAGT	7680
TTGCTCTTTG ATTTTCATTG AGTATAATAC AAAAATCCGA TGAACCTCAC CGGACTCTTT	7740
TATTTTGAAT TTTTGCTGC TTTACGCTTT TCAGCGATTT CGGCTGCCTT TCGAGGCAAG	7800
ACAATTTCCG TTATGTAAGC CGTCCCAAAA CGCAGTACAC CTGCAATAGG AGCAAAGACA	7860
ACTGCTAGAT AGTTATAGAA GAAATCGCCT TTGAAGGCAT AAGCTAGCGC TCCAATGATG	7920
AAAAATAGAA CGACTGCCTG AATCACTGCT AATAAAATTA CTCGTTTCAT GTGACCTCCT	7980
GACTCTATTA TAGCATGAGA ATCATCAAAA AGCCGACTAA ATTATTCAAA GCGTGAAGAG	8040
AAATACTGTA GACCAGACCT TTTCTGCTAA TGTAAGCCAA ACCCAAACCTA AAACCAAGGC	8100
TAAAATAGAC AAAAAATTGT TGCACATCAC CTGGAAAATG AATCAAGGCA AATAGAAGAC	8160
TAGATACCAG AAGAAAAATC AGGGTTCGTT TACTATTGTC CTGCTTAGGA AAGAGATAGC	8220
GTGCTAACAT CCCTCTAAAA ACAATCTCTT CCGTCAAAGG AGCAAAAATA ACCACAGCAA	8280
AGAATGAGAA AAGTGGTTGA GACAAGGTCA AGTCTGTGCG TATTTGCTGA TTTACTGAAG	8340
GATCATCTGG CAAGAAGAAT TGAACGACCA GAGATAAGAA CCAAACCAAG ACAGGAAGCC	8400
AAATAAATCG ATTAAGCCG CTCTTCTCAA TATGAACAGG AGCCTTCTGA TACCATTGT	8460
AAATGCCGTA CACATATACT CCAGCCAAGG CCACATAGAG TAGAGTAACA GCATAGGGTG	8520
AAGCGCTAA AGCAAGCGAC GCAGTCGCGA GCCCCTGAAT AAAGCCATAG ATAAATAAAA	8580

345

AGGATAGAAG GGCTAGAAGA ATCCAGCCAA GGTTTTAAAG TAATTTTCATA GATAACTCCT	8640
TTATTTGAAA TAACGTTTTA CCATAGGTAA CTGCATCACA TTGATATAAA CATGGATGGC	8700
TCCTACAAGC AAGAAAGCTA GTAACGAAT CTCTCCTGTC AAGAAAGAAA TGATAATAAG	8760
AAAAATATAT AAGGCTGGTA AGACATATTG GTGTAATTGG AATAAAATTC GAAAACTCTG	8820
TTCCAAATTA GCCTGACGCT CCCCTTCATC ATAAGAATTT ATATAGTTCA AGACATCCTT	8880
TGGTGTAGCG AAAAATTCCA AATCAAACG ACGAACAATC GCAATGGTTT TAAAAAGAGA	8940
TTTTTGAGCG ACTAAGAATA CCACAAAGAG TAAGAAAGAA AGGAAAAATG TTTGAGGGTT	9000
TGTATGCAAT ATAATCACCT CACTTAATGA AATAAAAATA GCCAATGGAA TCGCTACACC	9060
TGTAATATTA AAAGCAATGG TTCCAACTC AAGATTCCGA TACATTTGCA CATAATAGGT	9120
TTCAATCAGA TCGTCATCCA TTTCTCTGTG ATACAAAGAA TGAAATTTTC TGCTTTTCTT	9180
TAAGAAATG AAAGTCAAAA ACATACTAAT GAAACCTATC AGTAAACAAA TAGCTGATAT	9240
CCATGGCATC AAGGCTTTTA CATCTAAAAA AATTCCTGCG GATTGACAC GTGCCTTAAA	9300
CATCCCTACA AACATGCCCA AGAACCCCCC AAGACAATAG ACATCAAAAA TAACAATCTA	9360
CGTTTCTTTT TCATATTCAT TCTCCTTTT CACTTGCTAG ATTTTGGAT TTCTTTTCAA	9420
TCCATTCAAT TACTGGGATG AGAGCAAAGT AGACCCAAAC AAATGGGTCG CTTTGATAGG	9480
GATTAAACCA GCTTAGGTCC ATCCCAATCA GTAGAAATAC GCTGACTAAT AAAGCTATGA	9540
CCACTACATA ATAAATCACT TTATACTTGT TCATCACTCG TCCTCCTCCA AACGAAATAC	9600
CGATTGCACT GTTTCGTGA AAATTTGAGA TATTTTCAGG GCAATGATAA TGGATGGGGT	9660
GTAATCATCC CGTTCTAGTA GGCTAATGGT CTGTCTGGAA ACCCCTGCCA GTTTGGCTAG	9720
GTGCGTTTGA TTGAGACCAT CGCGAGCTCG AAGCTCTTTT AGACGATTTT TTAGTTGCAT	9780
GTTACACACC TACTCTCCGT CAAATTC AAC GGTTTGGATA TCCTCAATAC GTTGCAACTT	9840
GAATTTTCT TTTCCCGTAT TATCTACAG TCGTAGCTTT ACCCATTCCT CATCAACATC	9900
CACAACCTCC CAGTTATCTG GCCCAATATA CACTCCCGTT ATAATGGTT CCTTTCCAAT	9960
CATTTCTTGT AATAATCTCG ACATTTCTGC GTTTCCTTTC TCTTTTCGCT CAAGTCTTTT	10020
GATTTTATTC TCTAGTTTCT TGATTTTTTT AGAATTATTA GAATAAAGA AAATCATAAA	10080
TAGTATAAAT CCTAGTACCC ACATTATAAC TCCTTTCTGC TTCCTATTTT TTAACCTGAA	10140
TTCAATGTAA CATATCTTTT TCTTTTGTAC AAGTATAGTT GTCAAAAAA TTATGATTTT	10200
TGTCATTTTG CAAAAGAAA AGGTCAGGAG TAGGTTCTCG ACCACTTTAT CTATCATTA	10260
TACTCTTCTA AAATCTCTC AAACCACGTC AGCTTCACCT TGCCGTAGGT ATGGTTACTG	10320

346

ACTTCGTCAG TTTTCATCTAC AACCTCAAAA CCATGTTTGT AGCTGACTTC GTCAGTTCTA	10380
TCCACAACCT CAAAACCATG TTTTGAGCTG ACTTCGTCAG TTCTATCCAC AACCTCAAAA	10440
CCATGTTTGT AGCTGACTTC GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA	10500
CCTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTATTTG AGTATAAAAT CCTAGTTTGT	10560
CAAAGATTTC TGAGAAGTTT TGGCTGATTG TCTCAAGTGA CACTTGCACT TCTTCTCGGG	10620
TTTGGTTGTT CTTGACCGTC ACTTGTCCTG TTTGCACTTC GCTCTCTCCT AGGGTGATGA	10680
GGGTCTTAGC CGCAAAGACA TCGGCTGACT TGAAGTGAAG TTTTAGTTTA CGGTTGAGGT	10740
AATCAGCTC TGCTTTGAAA CCTTGTTGGC GAAGAGCCTG TACCAATTCC AAGGCCTTGA	10800
TATTTGCCCC TTCGCCAAG ACTGCGATAT AGACATCTAG GCGGTTTTCG ATAGGAGGG	10860
TCACACCTTG CTTTCAAGG ATGAGAAGCA GCGGCTCTAC ACCAAGTCCA AAACCAAATC	10920
CAGCAGTTTC AGGGCCTCCA AAGTAAGCAA CCAAACCATC GTAGCGACCA CCCGCACAGA	10980
CGGTCAGGTC ATTGCCCTCA ATCTCTGTGA TAAACTCGAA AATGGTGTGG TTGTAGTAGT	11040
CCAGACCAG CACCATATTG GTATCGATGA TGTAATCTAC TCCAAGATTT TCCAACATCT	11100
GACGCACAGC ATCAAAATGA GCTTGGCTTT CTTCATCAAG AAAGTCCAAG ATAGACGGCG	11160
CATTCTCTAC TGCCACCTTG TCTTCTTTTT CCTTAGAGTC CAAGACACGA AGAGGATTTT	11220
CCTCCAAGCG ACGTTGGCTA TCCTTAGACA AGGTCTCCTT GAGCGGTGTC AAATAGTCAA	11280
TCAAGGCTTG GCGGTAGGCT GCACGGCTCT CAGGATTTC AAGAGTGTG AGGTGCAATT	11340
TGACACCTTG AATACCGATT TCCTTCAAAA AATGGGCTGC CATAGCGATT GTTCCACAT	11400
CGGTAGCTGG ATTGCTAGAG CCAAACACT CAACACCAAT CTGGTGGAAT TGGCGCAAGC	11460
GCCCTGCCTG TGGACGCTCA TAACGGAACA TAGGTCCCAT GTAGTAGAAC TTGCTTGGCT	11520
TTTGCACTTC TGGGGCGAAA AGTTTATTTT CCACATAGGA ACGGACAACG GGTGCAGTTC	11580
CTTCTGGACG GAGGGTAATA TGACGGTCAC CCTTGTCTA AAAATCGTAC ATTTCTTGG	11640
TTACGATATC CGTTGTATCT CCGACAGAGC GACTGATAAC CTCGTAATGC TCAAAAATAG	11700
GCGTGCGCAC TTCTGCATAG TTGTAGCGTT TGAAAATCTC ACGGGCAAAG CCCTCAACGT	11760
ACTGCCACTT AGCAGACTCA GCAGGTAAAA TATCCTGCGT TCCTTTTGGT TTTTGTAATT	11820
TCATAGGGAA TCCTCTTTAA ACTTAATAGT CTTATTTTAC CATAAATAGA GGGATTAAAA	11880
CAGTAAGAAA AAAATTAGGA TTTAGATATC ATTTTGTAGA TTAAGAATTG TCAAAAAAAT	11940
AGCTAGCAAG GAAAGACCAA CAAATAGCAT CCAAGTCAAC TGTATATTCC ATACGGCTAC	12000
TAGTGAAAAA CAAGCTGTTC CCACAGGTAT GGATAAGGTA AACAATAGAC CTAAAAAATT	12060
ACTAGTACGA GCTAGAACCT CTGGAGCTAG ATTTTTCATG AGCATGGCAC TAATCTTTGG	12120

347

TTGAACTTTA CCAGACACAT ACAGAGTAAA GAAGAGAAAT AGCAAACCAA GCACGACTTG	12180
ATTGAATAAA TTAGCCAAAC CAACTAGACT AAGTCCTACG GTCTCCCACA TCATCAATCT	12240
AGGCAAGGAC TGCTTCCCAA AATAATCATT GCCCGTAAGG CTA CTGATGA TGA CTGATAC	12300
TAAAACACAG AATTGATTGA TAAATAGTGC CTCTGTATAA GAAAAATTCA AGAGAGAATG	12360
GCTCAAAAAG AAGATATTAT AAATTCACC CAAAGCGCCA CCCAAGGAAT TAATAAGCAA	12420
GACAGCAAAG AGCATAAAAC CAAAGTTTTT CTGTCCACTT TTAAGAAAAA CGAGACGTAA	12480
ATTTCGGTAA ATTGTTAGGA ACTGGTCTTT GATAGAAAGC TTCTCATTTT TTAAGTTTTT	12540
ACCATCAGCA GATGACATTG ACAGGCTCAA TTGCTTTTTT CCTAAAAAGA GGATAGTGGC	12600
TGATACTAGG AAAAAGCAGG CATTGATTCC CGCAACGAGA GAAAAATTGT TGACCGATAG	12660
AGCTAAGAGC CAGACTCCGA AAGCTTGACC ACCAATAGCT GAAATATAGG TGATGAACTG	12720
TGAAAAAGAA TAAGCCTCCA TCAGATCATC TTCAGTACT TTTTCCTTAA TAAGAGGCAT	12780
ACGCAGGCCA CCTGCAAAAT CACTGATGAT ATCACTAATG ACATTGATCA AACACAGGCT	12840
AGAAAAGGCA AAGAGACTAG CTTGCTGAAC AACTAGGGCT GCTAGAAAAA ATAGAACCGC	12900
CTGAAACAAA CCGCTATAGA CCATCCATTT GACCTTGTC CTGCTGTAAT CTGCCCGAAT	12960
CCCTGCAAAA ACTGTAAAGA GGGTCGGAAG AATCATGACA ATATTCGCCA TAGCAACAGC	13020
AAAAGATGCT TGTGACAAGG TCGATGCATA GACGATAAAG ACCAGGTTGA AAATCGAAAC	13080
ACCAAAAGCA TTGAAGAAGC GTGG	13104

(2) INFORMATION FOR SEQ ID NO: 35:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19250 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCGGGCAAAT AGTTTGAAC TTTTCATCAT TTTCTCCTTT AAAACTTTCT CTCCATTATA	60
GACTCTTTTC AGAAAGTTGT CAACAGAATT TTCAGAATTT TTGAAAATTA TTTTCAAAC	120
AACATCTTTG CAAAAAATAT GAATATCGTA AGCGCGTCAT AACAAGGTAT CTATCATTCA	180
TGGAGCTCCT CCTGTATACT ATTAGTAAAG TAAATATTGG AGGATATTTT AATGCCACAA	240
CCTATTGTTT CTGTAGAGAT TCCACAATCT CGTCGTTTTG ATTCTAAAAA GAGAAATGAT	300
ATTCTTCTTA AAATTCGTAT TGGCAAGCTT GAAGTAAGTT TTTTCAATC TCTCAATCTC	360

348

GAAATGATAG AACAGCTTTT GGATAAGGTG TTGCTCTATG ACAATTCATC TATCTAGCCT	420
AGGGCAGGTC TATCTCGTGT GTGGGAAAAC TGATATGAGA CAAGGAATCG ATTCAGTGGC	480
TTATCTCGTT AAAACCCACT TTGAATTGGA TCCTTTCTCC GGTCAAATCT TTCTCTTTTG	540
TGGTGGACGT AAAGACCGCT TTAAAGTCCT TTAAGTGGAT GGTCAGGAT TTTGGCTACT	600
ATATAAACGC TTTGAGAACG GCAGACTGAC TTGGCCCAGT ACAGAAAAGG ATGTCAAAGC	660
TCTCGCACCT GAACAAGTAG ATTGGCTGAT GAAAGGCTTT TCTATCACTC CAAAAATATA	720
GTAGATTGAA ACTAGAATAG TACACCTCTG CTTCTAAAAC ATTGTTAGAA ATCGATTTTA	780
CTGTCCTGAT CGATTTGTCC TGTATTATT TCATTTTACT ATAAATCCAT CAGAAAGTCG	840
TGATTTCTAT TGAAATGAGG ACTTTCTTTT TATACTCATC TGCTTTCAA AAGCACTCTA	900
GTCCATCTCC GATTAACGAT GGACTTTATC ACCTCCTTCT CCAGTCCTTG TATAACATCT	960
TGAAGTTGAT TCATGACATC TTCCAAAGTT GAAAGGCTT TATTCTTAAA TCCACGTTTA	1020
CGAATCTCTT TCCACACTTG TTCAATGGGG TTCATCTCTG GTGTGTATGG AGGAATAAAT	1080
GCAAAGCCAA TATTAGTCGG AATCTTTAAG GTACTTGATT TATGCCATAT AGCATTGTCC	1140
ATAACGAGTA AAAGATAATC ATCTGGATAA GCTTGTAAG GCTCCTATTC CTAAAGCCCC	1200
TTTATAACCT CTGCGAGAG AGACTATTGA CTCAGCCCTT ACTTCATGCG GATGAAACCT	1260
CCTATCGGGT TCTAGAGAGT GATAGCCATC TGACCTACTA TTGGACTTTT TTGTCAAGTA	1320
AAGCAGAGAA ACAAGGGATT ACGCTTTACC ACCATGATCA GTGTCAAGT GGTTCAGTAG	1380
TACAAGAATT CCTAGGAGAT TATTCTGGCT ATGTTTATTG TGATATGTTG CGGCAGTAAC	1440
TTAGGACTTT AGTCTCTAG TTCTGCCTAT GCGATAGCAG TCCAAGGTTT AGGAGTAAGG	1500
CGACGCTAAG CTGTTGTAAC TGCGAACAGC TAGAAGCTTA TCGTCAACTG GAAGAAGCTG	1560
CACTTGTGTTG ATGTTGGGCG CATGTGAGAA GGAAGTTTTT TGAAGTGCCC CCAAGCAAG	1620
CAGATAAATC ATCCTTAGGA GCTAAAGGTT TAGCCTATTG TGATCAGTTA TTTTCCTTGG	1680
AAAGAGACTG GGAGGCTTTG CCAGCTGATG AACGGCTACA GAAACGTCAA GAACATCTCC	1740
AACCCCTACT GGAAGACTTC TTTGCTTGGT GCCGTCGTCA GTCAGTTTAA TCGGGTTCAA	1800
AACTAGGAAG GGCAATTGAA TACAGCCTCA AGTATGAAGA AACCTTTAAG ACCATTTTAA	1860
AAGACGGACA TCTGGTCCTT TCCAATAATC TAGCTGAACG CGCCATTAAA TCATTGGTTA	1920
TGGGACGGAG TAAAAGAGTC CAGTGGACTC TTTTAGCCTA AGCTCAGTTT AAAAAACGA	1980
GGGTGGTTAT TTTTAAAAAA GCGAGGGTGG TTATTTTCTC AAAGTTTGA AGGAGCTAAA	2040
GCAAGAGCTA TTATTATGAG TTTGTTGGAA ACAGCTAAAC GTCATCAATT ATAGTGCCTT	2100
GAATCTATAA CAGTACGCAT CGACTGCTAA AATATTCTA TAAATCAATT TTCCTTTCCT	2160

349

AATCGATTG	TTCATATCTT	ATTACAATCC	ATTATAAATA	GCGAGAAATA	TCTATCCTAT	2220
CTTCTAGAAT	GTCTTCCAAA	CGAGGAAACT	CTCGTAAACA	AAGAGGTTTT	AGAGGCCTAT	2280
TTACCGTGGA	CTAAAGTTGT	ACAAGAAAAG	TGCAAAATAAG	AAATCTCCAG	ATTAGGAACT	2340
ATATATGAGT	TCTCTAGTCT	GGAGATTTTT	CAATAGACTT	CGTTATTGGG	CGGTACTTTT	2400
CGAAACTTTG	AAAACTTCAA	AAAACGGATT	TTTATCGCTC	TGAACATCAA	AAAAGAAAGG	2460
ACGAAATTTG	TCCTTTCTCA	AGCTTAGCTT	TTCTTCAACC	CACTACAGTT	GACAAAGAGC	2520
CCTTTATTCT	ATCAAACATG	AAGCGCAAAA	ACAAGCCAAA	AATCCGATAG	AATGGCTATC	2580
CCTCGACTAT	CAAGTAAGAC	ATTTCCATCA	AATACGTTCA	ATTTTACTCT	TGTTCTACTA	2640
AGAATTAATC	ATCTCGTTTT	GATTATTAA	AAATATACAA	TTCAGCTTTT	CCTCCAAACT	2700
ATTTTATCCA	CTATCCCTGT	ATAGCTCTGT	ATTATCTTAA	CAACTTTAGT	AGAGACATTT	2760
TCCTCAACAT	AATCCGGAAC	CGGTAATCCA	AAATCCTCAT	CTTGTGCCAA	GCTAACAGCA	2820
GTTTCAACTG	CTTGAAGAAG	AGAATTTTCA	TCAATGCCTG	CCAAAATAAA	TCCTGCCTTA	2880
TCTAAGGACT	CAGGACGTTT	TGTACTTGTA	CGAATACATA	CAGCGGGAAA	AGGATAACCT	2940
TGACTAGTAA	AGAACTACT	TTCTTCCGGT	AAAGTTCCCG	AATCAGATAC	TACAACAAAT	3000
GCATTTCATCT	GTAACAATT	ATAGTCATGG	AATCCTAGTG	GCTCATGCTG	AATCACACGT	3060
TTATCTAGTT	TAAAACCGCT	CTCTTGTAGC	CTTTTCTTTG	ATCTAGGATG	GCAAGAATAT	3120
AAGATTGGCA	TATTATACTT	TTCAGCTAAT	TGATTAATTG	CTGTAAAGAG	AGAAATAAAA	3180
TTTTTATCTG	TATCAATATT	TTCCTCACGG	TGAGCTGAAA	GTAAGATATA	ACCTCCTTTT	3240
TTCAATCCCA	AACGTTTCATG	GATATCTGAA	GACTCAATAG	CAGATAAATT	TTTATGTAAAC	3300
ACTTCTGCCA	TAGGAGAACC	AGTTACATAT	GTGCGCTCTT	TAGGTAAACC	ACACTCATGT	3360
AAATACTTAC	GTGCATGTTT	AGAGTATGCT	AAGTTAACAT	CTGAAATAAC	ATCAACAATC	3420
CGACGATTAG	TCTCTTCCGG	TAGGCACTCA	TCTTTACAGC	GATTGCCAGC	CTCCATATGA	3480
AAAATTGGAA	TATGTAAACG	CTTGGCAGCA	ATAGCTGATA	AACAAGAATT	TGTATCCCCT	3540
AAAATCAATA	AAGCATCTGG	TTTAATTGTA	TTCATCAATT	TGTATGAAGT	ATTAATAATA	3600
TTCCCTACAG	TAGCACCAAG	ATCATCTCCA	ACAGCATCCA	TGTATACGTC	CGGAGTGTCT	3660
AACCCTAAAT	TATCAAAGAA	AATACCATTT	AAATTGTAAT	CATAGTTTGT	TCCAGTATGT	3720
GCCAAAATAA	CATCAAAATA	CTTTCGACAT	TTAGTGATAA	CACTACTTAG	ACGTATAATC	3780
TCGTGACGTG	TTCCACAAT	AATCAATAAC	TTAAGTTTGC	CATTATCTTT	AAAGTGAATA	3840
TCACTATAAT	CTGTCTTAAT	TTTCATTTAT	TTCTCCACTT	GTTCAAAAAA	AGTATCTGGA	3900

350
TGTCTAGGAT CAAATGACTC ATTAGCCAC ATGACAGTAA TTAGATTTTC TGTATCAGAA 3960
AGATTAATAA TATTATGTGC ATAGCCCGGT ATCATATGTA TTGCTTCAAT CTTATCGCCC 4020
GACACTTCAA AGTTCAGAAT AGGATACTCT TGACCGTTTT CATCCAGCCC TATCCTACGC 4080
TCTTGATTA AAGCACGACC AGAAACAACC ATGAAAAATT CCCACTTAGA ATGATGCCAA 4140
TGTGCCCCCTT TGGTAATGCC AGGTTTAGAA ATATTAACAG AAAATGACC CGTATTTTCT 4200
GTTTTAATA ATTCCGTAAA ACTACCTCGT TCATCTATAT TCATTTTTCAG AGGAACTTA 4260
AACTTATCTA CTGGTAAATA AGATAGGTAG GTAGAATACA ATTTCTTTT AAACGATCCC 4320
TGAGGAATTT CAGGCATAAC TAACTATCA GGCTGTTTT TAAATGTTTC TAATGAGAG 4380
ACAATCTCTC CTAAGGTGTC ACGATGAGTC GTTGGTACGT AGCAGTAGTT TCCTGATGGG 4440
CTAGGTAAGA TTTGTAATCC ATCTAGATTA CAACGATGAG GATTTCTTTC CAATGCAGTT 4500
AGACACTCTT GTATCAAATC ATCAATATAC AGCAACTCCA ATTCTACACT TGGATCATTT 4560
ACTTGAATAG GTAAATCGTG AGCTAGATTA TAACAGAAAG TTGCTACAGC AGAATTGTAG 4620
TTAGGACGGC ACCACTTCCC ATAAAGATTC GGGAAACCGT AAATAAGAC AGGTGCTCCC 4680
GTTTTCTTC CATATTCAA GAAGAGTTCT TCCCCTGCTA GCTTAGATTG TCCATATATA 4740
GAGTTTGAAA ATCGGCCTTC TAACTAGCT TGAGTAGAAC TTGAGAGTAG AACAGGACAA 4800
GTGTTTTTCAT ACTTTTCTAA AATCTCCAAT AATCTACTTG AAAAACCCTA ATTTCCCTCC 4860
ATGAATTCAT CAGGATCTCG TGGACGATG ACACCAGCTA AATGGAATAC GAAATCGGCC 4920
TTCTTACAAT ATTCATCTAA TAAATCGGA TCTGTATCAC GATCATACTG AAAATCTCT 4980
CCAATCTCTA AATTAGGACG AGTCCTATCT CGTCCATCTT TCAAAGCTTC CAGAGTACAG 5040
ATAAGATTTT TTCCTACAAA TCCTTTCGCT CCTGTGATTA AAATATTTT AATCATGCCC 5100
CCTCCTTATT TTATATGCTG TTTAATAGT TAACTCTCTC GACAATACAT GATACATTAT 5160
ATATCCTTGA TAATTTTAAT GTATCTTAAA AGATTTTACA TCTCTCGTC TGCTACCATA 5220
TCACGAATTG CTGCTGTAT TTCATCTAAT TCTAGCAACT TTCTTTTAAAC TTGCTCTACA 5280
TCCATCAAAT CGGTATTATT ACTATTGAAT TCTGTCAACA AATTTCTATT CGTACTACCA 5340
TCTTTGAAAT ACTTATCATA GTTAAGATTA CGATTATCAC TAGGAATCT ATAAAAATCA 5400
CCCAAATCAA TTGCATTTGC GCACTCTTCG TTAGTTAATA GTGTTTCATA CCTTTTCT 5460
CCGTGTCTAA TACCTATAAT CTTAATATCT TGTCTGAGG CAAAAATTC TGATACAGCC 5520
TTAGCCAACA CTTCAATCGT ACATGCTGGT GCTTCTGAA CTAGTATATC TCCAGATTT 5580
CCTTCTTCAA ATGCAAATAA AACCAAGTCT ACTGCTTCTT CCAATGTCAT CACAAAACGT 5640
GTCATGCTAG GTTCAGTAAT TGTAAAGACA TTTCCTTGCT TAATTTGCTC AATCCAAAGA 5700

GGAACGACAG ATCCACGGCT ACACAGAACA TTCCCATAGC GAGTCACACA TATCTTTGTA 5760
TGCTCAGGAT TTACCGTCCT GGACTTAGCA ACAGCAATCT TTTCCATCAT AGCCTTGGAT 5820
GTTCCCATAG CATTGACAGG ATAAGCCGCC TTATCTGTAG AAAGACAGAT AACTTGCTTT 5880
ACACCAGCTT CGATAGCCGC AGTGAGGACA TTCTCCGTTT CCAAAATGTT AGTTTTTACC 5940
GCTTCTACAG GGAAAAATTC ACAAGAAGGT ACTTGTTTAA GAGCAGCAGC GTGAAAAACA 6000
TAATCCACAC CATGCATAGC ATTTTACC GAAGCTAAGT CACGCACATC TCCAAGGTAA 6060
AAACGGATTT TCCCAGCCAC TTCTGGTACT TTTACCTGAA ACTCATGACG CATATCATCT 6120
TGTTCTTTT CATCTCGCGA AAATATACGA ATCTCTGAGA CATCTGTTTC TAAAAACGC 6180
TTGAGAACCG CATTCCTCAA TGAACCTGTC CCTCCTGTAA TTAGGAGAGT TTTTCTGTA 6240
AATGTGACA TATATTACAC TTCTCCTCT AGTATGTCTG CAATTTCTT ACAAGCCGTT 6300
CCATCTCCAT ATGGATTGTA AGCTTGACTC ATTGCTTGAT AACTGAATC ATTTCTAAT 6360
AATCTTTAA AATGCCTATA AATATTATT TCATCAGCAC CTACAAGTT CAAAGTCCCT 6420
GCTTCAATTC CCTCTGGACG TTCAGTTGTA TCTCTCATA CCAAACAGG TTTCTCTAAA 6480
CTTGGAGCCT CTCTCTGAAT ACCACCACTA TCTGTTAAAA TTAAATAACT TCTTGATAAA 6540
AAATGTGAA AATCTAATAC TTCTAAAGGT TCGATCATCT TGATACGTT ACAGCCACTT 6600
AGTTCTTCCT CAGCAATTTG GCGAACACGA GGATTCATAT GGATAGGATA AATAGCCTTG 6660
ACATCTGAAT ATTCTTCAAT AATCCTTCTA ATTGCTCTAA ACATATGTCT CATCGGTTCA 6720
CCAAGATTT CACGACGATG AGCTGTAATT AGAATAAACC TGCTTTCTCC TATCCATTCT 6780
AACTCAGGAT GCGTATAGTC CTCTTGAATT GTAGTTTGTA AAGCATCAAT CGCCGTATTA 6840
CCTGTCACAA ATATGCTCTC TGGAGTTTTT CCTTCTCTTA AAAGATTATC TTTTGAAAGT 6900
TGTGTTGGTG TAAAATGATA CTGAGCCAAA ACCCCAACCTG CTTGACGATT AACTCTTCA 6960
GGATATGGTG AATAGATATC GTAAGTGCGC AAACGAGCTT CAACATGACC AATTGGAATC 7020
TGTAATAAAA AGGCCGCCAG TGAAGTAGCG AAGGTCGTAC TTGTATCCCC ATGAACTAAC 7080
ACCAAATCAG GTTTTTCTGA CTCTAAAATA GCCTTCATTC CTTCCTAAAAT GCCAATGGTC 7140
ACATCAAATA AAGTTTGTAT ATCTTTCATA ATAGACAAAT CAAAATCGGG AATAATCCCA 7200
AATGTGTCCA AGACCTGATC CAACATTGTA CGGTGTTGGC CCGTAACGCA AACTAATGTT 7260
TCAATATTCT TACGTGTTCT TAACTCTTTG ACCAAAGGAC ACATCTTGAT GGCTTCTGGA 7320
CGAGTTCCAA ATACTACAAC TACTTTTTTC ATATATTTAC TTAATCCTAA CAAATAATGA 7380
ACGGTTCTTA AAATAAATTA GATAACGGCT AATCCATAAC ACCACCTCAG ACATACTTGA 7440

352

ACAAATAGCT AATGTTACTA AACTAAAATT ATCAGACAAG ATAAATATTC CTAATCCCAA	7500
AGTTTGGACA ATCGAAGCTA ATATAGTTGT CATTGTAGTT TCTTTCACCT TATCAATAGC	7560
TCCTAAGACA GGCCATCCGT AAATCATAGA ATAAAACTA GCAACAAAAG CGGTAATAA	7620
GTACTIONAAGA AAATCTGCTG AAACGGTATA TTTTTCACCA CCAATTATAG AAAGAATTG	7680
ATTTGAAAAG AATAAACTA TCAAACTCC AAAGATAATA GGAATAAACA TAATCCGATT	7740
AATACTCTTA ACCGATTGTA TATCTTTAGT ACGTATCATA TGCGGATATA AACTATTCGC	7800
TATAGGATTA TACAATGATT TTGCTGCTGA AAGCAGTTGC ATTGCTATCC CCCAAAAGGC	7860
TATCTCTTGA CTTTGTAAT AAAAACC GA AATGACTGTC GTAAAGACGC CAAAAATAGT	7920
AGTTGCAAAA TTGGATAAAA AATAAATAGA GGATTCCTTT AAATCTTTAA CCCAAACAGA	7980
CAGATAAGAA AATGATAATT TAATCCATA ATAATGAAGG AATCTATAAG AACTACTGC	8040
AGCAACTAAA TTCCCAATTC CTTCCAATAT AGGAATCCAT AAAATAGAAG AATCATCTTT	8100
TACTACAATA AATGTCAAAA TTGTAATGAT AGTTT TAGAA ATAATATAAG GAATTGCAAC	8160
TGCATGCATC TTTTCAATTC CACGAATAA AAAGTCAAAG ATAAAAATAT TGGTCACTGT	8220
AGCTAACAAA TAAAAAATG AAAAAAGAAT ATTCTCTCTC ATTATGCGA TTGCCACAT	8280
CAATATGGTG TAAATTAGAA TCGAAATGAT AGATAAAAAAT ATTTTTCAT CTAAGATATC	8340
TCCAATATC CTTCCAATCT TTGAGGGAGT AGTACAAGCA TTTACAATAT TTTTGTAGC	8400
TGATATCATG AAACCAAAAT CAATCACCAG TTGAACATAA GCTATTACG CTTTAACATA	8460
AATAACCATT CCATACCGGT CTAGCGAAAG CACCCTTGTC AAATACGGGA GTGTTAATAA	8520
AGGAAATAGT AATTTAACAA TATTGAGAT ATAGAGAGAA CTTGTATTTT TTATAAATGA	8580
AATTCTATCA ACTTTCACGA ACTAGTCCTT CCAAAAAAAG ATCTAAATAG TCCAAACTAC	8640
TTCTCGCTTT CAACACCAAT TCTGAAGGTA TTGTTATCGG TTTTAGATGA AAAGTTTCAA	8700
GTTTCTTTAC AATACTATTA ACACTTGAAT CAAATAAAGA TTCACAACGT TGTAACCTC	8760
CAATTGCTCC ATAATAACGT GCTGTTTTT CTGGATGGCA TGCAATGGCA ATCAGAGAT	8820
TATTAAAAA TGTTGCCACT ACCCAACAT GTAATTTACA AGTTAAAACC ACATCTACCA	8880
TTTTCAACAA TGATGTCATT TCTGCAGGAG AATGATACTT GAATTGAAAA CAATCCTCAG	8940
TTCTAACTAA TTTTCTAAT TCCTGATAAT AAGCATCTC ATAAGGTAGA ATGGAATCCG	9000
AAGTTACTAC AACATAATAG TTAGGATTGT TTTCTAGAAA AAGACTAATT GATTCCGCAA	9060
ATTTTTCAG AGCTTTTTT GAATGATTAT AGTGAACAAG AATTATCTTC TTATCTTTAG	9120
CTTCTCTTTT CAATTGACAC AGCTGCTCTG TTTTCTCTC TCTTAATTTA CTTGAAATAA	9180
TTAAATCAAA GGTTCATGC ACTGGAGCCG AAGGCGACAA ATGCTTCAAA GAATCAAATG	9240

353

ATTCTCGATC ACGAACTGTA ATAAATTGAG CATGATTAAAT AATTCTCTTT ATACCATAAT	9300
TCATCAAAGA ATCGTTATTA GGCCCTGCAC CAATACCTAA TACTCCTATA GGCTTTTAA	9360
AATATGAAGC CCAAATTCCT AAAGGTAAAA ATCGTTTAAA TTGGATTAAA TTATCACGAA	9420
AACGTGCATT ATGCCCTTCC CAAAATATC CTCCCCGGAT ATACAAAATA GCATCTGCTT	9480
GTTTTTAGT AAAACTTTGT TTTTGGCGAT ATTCTTTCAA GTACATTGA AAGAAATCTG	9540
ATGGATTATA AAAAGAACT TCATATCCTT TAGATTCTAA TAAATCATAG ACAATCTCAC	9600
CGTAAAGATA ATCACCATAA TTACTTGAAC CATAATCCGT TGCACCATGT AACATAATTT	9660
TTTTCACCAC TATTTTTTCA ACCTCTAAA AATAAATATC ATAATCAAAC TATACATAAT	9720
AGGACGATAA ACATCTATTG AACTACTTCT CACTAAAAGC AATAGTTGAG AAATTACCGA	9780
AAAATAAATA ACTTTTGAGA TTTTACTTGT TTGAAAAGCT CTGAAATTTA ATCGCCATCC	9840
ACTAAATATT CCCAAAACAA AACTCCAAAA AACACCACCA TAGTAACCAA AGTTCCAAAA	9900
TAATCTTCC ACAAAGAAG AGCCTACAGG TAACCCCAA AATTTATTAA TAACAACCGT	9960
CGTGATGCT TTATCAAAAA AATCACCAC TAACCATCCA ATAGGAAAAA TTGATAGGAT	10020
AGTGCGTAGA AATGTCATCC CATATTCATA TGGAATGCTA CTAGGCACAA CAGTTACAGC	10080
AGAAGCTACT GTTAGGCTGG TCAGTCCGA CTCTGAAAT ACTTCCCCTA GTATATTCTT	10140
TACAAAATCT AATGAAGAA AGGAATCAA TAAGTATATA CCTATAGTAT TCAAGTCGAA	10200
ACGGTGCCCC CTAATAACAA CTAATACATT TAATAGAAAT ACAGTTACTA TTAAAAATAC	10260
AAGTACTCTT TTCTTCGAAA AAGTAATCCC TAAAGATTGT GTGTATACTA AAACCAACGC	10320
CAAGATTGAA AACACCTGGA TTTTACGACT TCCTGTTAGG ATCATTATCA AAATTAGGTA	10380
AAACAACATT ACCCAAAAA TAGTACGCTT TATAACTCGG GACAGCTTAT CTGAATAAAA	10440
CAAGGAGAAC ACACCAGGAA GCATAAGTAC TCCTAAATCA TCTATTATTC CTGAAC TAGC	10500
TGCCTCTGAA TATGCTGAAT AGCTATTCGC CGCTCTAACT GCTAGTACTG TTTTAGAATC	10560
AGTTATTACC CTAGAAATAA AGCCCACTCC TGTTAAAATC CTACCCGCAT TGTACAAAAT	10620
TTTCTCTTCA TTTTCTGAT AATTTGTAC TTCTGAATGA TAATGTACCT TTCCATCACT	10680
ATAAAAAAAT AAATAGCCTA CAGAATAACA AAACAAAATC CAAATTATAA AAATATATGA	10740
ATGAAATAAT TCTTCATTAT TATAGAAGTT ACTAGGGCTC CACAGCAGAG TTGTTTGAAA	10800
CCCCATATAC TCATTGAAAA TTAATCCAAA CATAAAAAAA TAAGATAAAA TCAGATACCA	10860
TACAGAAAA TCATATATAC TAACTTTTGT TAAAATAAAA CCAGTAATTT GAAAAATAAT	10920
TAGAAAGCAA ACCCATATAA ATATAGACGG AACATAATTA GATATAAGAA AACCATTATT	10980

354

CCAATTATCG AGAGTCCAGA ACAAGTAACA GAAAGCAAAT ATAAACTTA ATGTCACTAG	11040
TGTCACCTCA CAAATATACT TTGTCTGCAT CTATATCTCC TTTATTACAC ACATTCTTG	11100
ATAACGATTC AATAATTTAC TAGCTTGATA ACAAATATCA TAGAGTCCAT CTGTCATACT	11160
GTTATTTATT TCAAAACGAT TGCATTCTC AGATGTTAAA GACAGTACTT TATCTTTCCA	11220
TAGCAACACA GACTCTTCGT TGATAGGTAA GTAACATAAG TTTTGGTCA CATCTACTTC	11280
TTGCGTCACT GTATCTGACG ATAAATTTG TAATCCCGAT GCCTGAGCCT CTACTAGAGA	11340
AACAGGCAAC CCCTCATATT TAGACGGAAG CAAAAAACA TCCATCGCAG ATAATAAATC	11400
AGAAATATCA GTCCTTCTCC CTAATAATAG CACATATGGG GTCAGATTTA GTTCTAAAGC	11460
TTTCTGTTTT AATTTCTGCT CATCCTCACC ATTACCAACT AGGAGTAAAA TAACATTTGG	11520
TTTGATTAAA ATGAGTTCTT TTAAACGTT AAATAAATAA CTTTGGTTTT TTTGATCTGA	11580
TAGGCGAGCT ATATTTCTTA ATACGAACTT ATTTGACACA TCTAATCTC TACGACATTT	11640
TTCTCTAACA TCTGACAAAA ATTGATACTT TTTCAAATCA ATTGCAATTA AAATAATTC	11700
AATTTTTCCG TCTTTATACG CTTTCTCTCC ATATAACCAC TTAGCCGAAT CTTCCCCACA	11760
TGCAAAACCA TGAGTTGCTA AGATTTTAC CAAATTTGT ACTAATTTAC GCAATACTTT	11820
TTGAAACTG TTTCTGTGA CATAAGCCAT ATGACTATGA ATAATTTCTA TTTTACAACC	11880
AATTATTTTA GATAAGATCA GACCAATTGC AGATTTATAG CCATGGCAAT GAACTATATC	11940
ATAATCTCCT TTCTTTATTA TTCTAGCAAG AGAGAGAAAC TGATGTAGAG GCTTTTTCTT	12000
TAATAGAGGC ACATGATAAA CCTTTGCACC CAATTCTTTC ATTTTATCCT CTAAAAATCC	12060
TTGTCTTTTT CCAGGCACAA TAAATCAAA TTGAATTTTT TTTCTATCAA TGTGAGAATA	12120
ATAGTTGAAT AGAAAACTTT CTAATCCACC ACTATCTAGT GTTGTAATA GATGTAATAC	12180
TTTAATCATT CTCTTCTCTT AAGCTTAAGA TTCGCTTCTC TAATCTATT TCTGTTTTTT	12240
GTTTTCTTAA ACTAATCTG TCCATGAAGT TATCACAAT CTTAATTAGC TGTTCCTGT	12300
CAAGGTTTTG AATATACAAA GCCAAACAAT CTTTTCCGA TTCATCCTTC ATAGTAAAA	12360
CGAAACCAAA ACCATCTCTT ATTGACACTT TTTCCATATA AGTATCTTCA CAACTAAAA	12420
TAGGTTTATA CAACAATGCA GCAAAGTAGA GTTTATTAGA CAAAGCATAG TCTAGTAAGG	12480
GAGTGTGATT CCCGTATAAA TTCAAACAA CATCTGTATT CTTATAAAAA GACATGGTAT	12540
CTTTAGGCTG GAATGTGTCC ACCAAGTTAA CATGTCTGAT ATTTTTTCT TGACAAAAT	12600
CCCTTAATTC TCCTGCATTA GTACCTATAA AATTCAACTG AAATCGACTG TCATTTGCAA	12660
AAAAATCGAT TATTTTTTTA TTTTGTCTT GAAAACGAAT TAAACCAATG TAGGAAAGTT	12720
GAATTGGAAG CGTACTATTA TTTTTTAACT GCTTTACCTC GTTTAATCTC ATCATATTGG	12780

355

GTAGGTTATG GGTAGTAAAA TACTCTCCCA TTGGTAAAAA AAATTTATAG CCGTCTGAAG	12840
AAACGATATT CATTAAGAA TTTTTCACCA ATTGTTTCTG AACCAAACGA TAAACCAAAA	12900
ATTTTTCATA ACTGTAATCA CGAATATCAT AAATATATCT ATTTTAAAT GAAAAGAGAA	12960
GAAAATCTAC TAAAATGAAA GACACAATAC TATGTAACGG CAATATCATA TCATAATCAT	13020
TTTCTTTTAG CTTCTTTTTA ATTTCTTTTC TGAATTTTAC ATAACCTAAT ATCTTACTTA	13080
ATTTTCCTTT ACCAGAAAA GAAATACGAT AGTAGTTTGT TTTTGTAAATA ATCTCGTTAA	13140
TATTCTTATC CCAATATATA ACATCGTAAC TAATAGACAG TTTCTTCAAT AATTCCTTAT	13200
AAAAATTGAA GTAAGGAGTT AGATATATAT TATCAGATAG TATAAACAGT ACTCTCATT	13260
AATTATCTT TCTTACTTTC CCTCTCTAAA CATGTCTCCA GTTCGAGCAT AAACGTCTCT	13320
TTTGAAAAGT GATTTTCATA GTAACAACGA GCTTCTTTTC CTAACCTCTCT TTGTCTCTTA	13380
ATAGATAACA TACTAAATTT ACAAATATTT TTTGCCAATT GTTTTACATC TCGTTCGGGA	13440
CTAACATATC CACAATTGTC TTCTTCTACA ATTATTTTAG CATCTCCTGA AATGACACCT	13500
ATAATTGGTT TGCCTGCCGC CATATAAGAG TGTACCTTCC CAGGTATAGT ACGAGAACT	13560
ATCGAGTCTC CTATTAAAGA AACTAACATA GCATCTGATT TTTTATAGAA GGATGGCATT	13620
TCCTCCAAAG AACGTCTTCC ATAGAAGGAA ATATTCTTTA ACTCCAATTC ATGAGCTAAT	13680
GCTTTCATGC TTAACAATTC CGTACCATCT CCAACAAAAT GAAAATGAAT TTTCTTGGGT	13740
AAATTGGTAT TCTTCTCTAT CAAACTGGCA GCTTTCAAA TAGTTTCCAA ATTTTGTGCT	13800
TTGCCAATAT TACCAGCAAA AGTTAGGTCA AACTTTCTT TATTAATAT AGATTCTATCA	13860
GGGATAAAAA GATCTTCTGC ATATTGTGGC AAATATGTAA TCTTTTGTTC GGATATGTCA	13920
AATTGCTTCA CAAAATAATT TTAAATGAT GGACTAGTGA CAAATATATA ATCACTAGCT	13980
CGGTAAACTT TTTTGTAGAT AAATTTAAAC AGCTTGAAAA TCAAGCCATC TTGTTTCACT	14040
CCACCTACGG TTAAACTATC TGGCCAAACA TCCATACAAT ATAGAAACAT CGGTTTCTTA	14100
TATTTTTTTT TATAAGCCAT ACCAGCCCAT GCCATCATAA CTGGAGACAA TTGGTTAACG	14160
AATACACAGT CAAAATTCGA TCCATCTTTC GTTTTATACC TCCCAATAA AACTCCTAAA	14220
GTAGAACTAA TTGCAAGCT AAAATAATTC AACAAATCGAA ATACAACACT TTTTTTCTA	14280
GGGATTGTAT AAGAACGATA TATCGTAACA CCTTCTATA TCTCACGTCT TTTTTATTA	14340
TGACGATAAT CTGCATATAT CTTCCCTTCA GGGTAATTAG GAATCCCAGC CAAAACAGAG	14400
ACTTCATGCC CTTTTCGAAC TAAATCTTCA CAAATATCTG ACAACCTGAA TGGTCTGGC	14460
TTATAATGTT GGCAACAAA TAGTATTTTC ATTGTCCAAT TTAACCTTCT TTCTTACCAC	14520

356

TACCCCTCTAC AATACCTTTT CGTTTCAGTA CGTAAGGTAT TGTCTTAACT ATACATCTAA	14580
TATCCATTAT CAAAGACAGA TGTTTAACAT AGTAGCCATC TAACTCCGTC TTCATCTCAA	14640
CAGACAAAAGT ATCACGCCCG TTAATTTGTG CCCATCCAGT TAACCTGGC AAGATATCAT	14700
TTGCTCCATA CTTATCTCTC TCTGCAATCA AATCTAGTTC ATTTATACCC GCTGGTCTAG	14760
GACCTACAAT ACTCATATTA CCAACAAGAA TATTAAACAA TTGTGGTAGT TCATCCAAAG	14820
ATGTTTTTCG CAAGAAAGCC CCTACTTTTG TAATCyATTG CTCTGGATTA TATAAGTTTC	14880
GAGGCGCCAC ATTTTtaggt GCATCTATTT TCATAGACCT AAATTTCAAA ATATAGAAGT	14940
ATTCTTTATG AATACCAAAG CGTTTTTGCT TAAATATAAC CGGACCTTCT GAATCAAGTT	15000
TAATCGCAAT TGCAATTATC ATAAAAACCG GACACAATAT TATTATCCCT ATTAAAGATA	15060
ATAATATATC ACCTAATCGT TTTATTATAC CGTACATAAA CAACCTCCAA CTATAAATTC	15120
TATTTCCATT TTTCAATCTA TTTCCATTTG ACAAATTAAA TCAGGCAGTA CATGCAACTA	15180
CAGAAACTCA ATATATATTT GGTCACTCAA TGATTTTCAG AAATATAATT CTTTATCCT	15240
CTACGTCAGA TAAAACTTTT CTCCATCTAA ACAAATTTA TTTGTTTCAG TAATATATGA	15300
GTTCTCAATA ATGAATTAGA AGGTCCAGTT CAATTATTCT TCCAAATAGA CCGAATATTA	15360
TTTGAAGACA TATCGGTTTC TGAAATTGCA ATCAGTACAT AAGCTAATAA ACTGATAAGT	15420
ATGCTCTGTA AGAATGCCAG AGTTATATTG TAGTCCCCTT CCATACTATA TTCATTTTAT	15480
TTTTTACCAT AATTTCATA GGAACCGTAA ACTCCATACT TATTAACCGA GATATCCAAT	15540
TTATTTAAAA CAACTCCTAG GAACAGTTTC CCGTTTGTG TTAATTGTTG TTTCGCTTTT	15600
TGGATATCAC GTTATTCGC CTCACCTGTT GCTGTTACCA AGATGGACGC ATCACACTTT	15660
TGAGTGATAA TTGCCGCATC AATAACAATT CCAATAGGCG GTGTATCAAT AATGATATAA	15720
TCAAAATATT TACGCAATGT TTCAATCATA TCATTAAAAT TTTTACTTTG TAACAAGGCT	15780
GTAGGGTTTG GTGATACAGA TCCCGATTGA ACTACAAATA AATTTTCAAT ATTTGTATCA	15840
CATAAACCGT GAGATAAATC AGCTGTCCCA GATAAAAATT CTGTAGCCC TGTAAATTTT	15900
TCACGAGATT TAAAACTCC TAACATAACT GAATTTGAG TATCGCCATC GATCAAAAGA	15960
GTTTTATAGC CTGCACGCGC AAACGACCAT GCTATATTTA TGGAAGTAGT TGTTTTTCCT	16020
TCCCGAGGGT TAACAGAAGT AACGGAAATT ACTTTTAGTT TATCTCCGCT CAACTGTATA	16080
TTTGTACACA AGGCATTGTA ATATTCTTCT GCCTTCTTAA TGAAGTCCAG TTTTTTTTGT	16140
GCTATTTCTA ATGTCGGCAT CCTTCTCTCC TATTTCAACT TACCCAAGTT TGGCACAAT	16200
CCCAAAAGTG TCATCTGCAA TGTATTTTCG ATATCTTCCG GACGTTTCAC ACGAGTATCC	16260
AAAAGTTCAA GATGAAGAAC TATAACACTA GTTCCAATCA CCCCTGCCAA AAAACCAATT	16320

357

AGTGTATTGC GTTAAATATT TGGCGAAGAC GGGGATATCG CCGGCCTTGC CTCCTCCAGT	16380
GTGTGCACGT CAGAAACACG AGTAATACTG ATAATTTTTT GAGCAGCTAC TTCTCTCAAA	16440
GAGTTAGCGA TACGGCTTGC CTCTTCAGGA ACTCGATCAT TAACTGAAAT AGAGACAATA	16500
CGGGTATCAA CTGGTACTGT CACTTTAATT TTATTAGCCA AACCTTTTGG CGTCAAATCT	16560
AGTTTCAAAT CAGAAACAAC TTCCTCCAAA ACATCCTGCG AAAGGATAAT CTCACGGTAG	16620
TCTTTTACCA GATAAGTTCC TGCCTGCAAA TCCTGATTG TCAACCCCGG CTGTCTCCT	16680
TGATTGCGAT TCACTACGTA AATTCGCGTG GTACTCGTAT ATTCTGGCTT AACATAAAAA	16740
GTGCTATATG CAAAAGCCCC CGCACCTGTC ACAAGTGCCA CTATTAAAAT CATTAGCTTG	16800
CGTTTCCACA AGCTTTTAAC TAATGAAAT ACATCGATTT CTATCGTATT TTGTTCTTTC	16860
ATCATTCTC CTAAATTAGT TGATCCATTA CAATTTTTCG AGGATTGTCT ATAAAAAGTT	16920
CCTGAGCCTT CGCTTCTCCG TATTTTGGG TAACAAGGC ATATGCTTCT GCCATATGAG	16980
GAGGTCTACC GTCTAGATTG TGCATATCAC TTGCAATGAC ATGAACCAA TCCTGCTCTA	17040
AAAAATACTG AGCTCTTTT TTCTATGAAT TATAACGTTT GCCAAAAAGT TTGGGTTTGA	17100
GGACATGTGA ACTATTTACT TGCGTGTAA AGCCCATATC GATCAGTTCT CGAACGCGTT	17160
TTTCATTATT TTCAAGAGCA TCATAGCGCT CAATGTGGGC AATGACTGGA GTAATTCCCA	17220
ACATCAAGAT CTGCTCAAG GCGCTATGAA TATCGCGATA AGGAGTGTTC ATACTAAACT	17280
CTATCAAGGC ATAACGACTA TCATTGAGGG TCGGAATCCG CTTTTTTTCC AGCTTATCCA	17340
GAACATCTGG TGTGTAATAA ATTTAGCCC CGTAAGCAAT GACCAAGTCA CTCGCCACTT	17400
CCTTAGCTAT TTCCCGAACC TGAAGAAAGT TTCTGCTAT CTCTCTTCC GGAGTTTCAA	17460
ACATGCCCTT GCGACGGTGA GAGGTAGAAA CAATGGTTCG CACCCCTGT CTGTAGGATT	17520
CTGCCAAGAG AGCCTTGCTT TCCTCTCTG ACTTGGGACC GTCATCTACA TCAAAAACGA	17580
TATGCGAATG GATGTCTATC ATTTATCTA CCCTCCATCA CATCCTGTAT AGCTGCTTTA	17640
ACTACAGCTA AACTACTATC ATCTATTTC ATCACAAGA GGTACTGTC TGGCATTGCA	17700
TAAGAAGGAA GATCCATCCG ACCTGTCCCT TTTAAATCTT GAGAAATTAC TTTATAATTC	17760
CCTCCACTTT CTAAGTATG ATTGACCAA TTTATCATGG TCTCAAGTGG CATATTTGTT	17820
TGGATAGAAT CTTGCAAGCT ATTAATGATC GTACTATAAT TTTTCAGCAC TTCGGTTGAC	17880
GTTAATTTTT GAAGGATAGC CACAATCACC TTTTGTGAT GCGCCCCGCG GTCACGATCG	17940
CCATCTGCTA GGGAGTAGCG CTCACGAACA AAACCGAGAG CCTGTTCTGA ATCAAGATGA	18000
ACATTGCCTG CAGGTAATA CTTTCCATTC GTATGGGACG TAAATCTTG ATCATTATAA	18060

358

ACATCAATTC CACCCAACAA ATCAATCAAT TTCAAAAACG AAGTGAAGTT CAATCGCACA	18120
TAGTAATTGA TATCCACTCC ATAGAGATTT TCTAAGGTGT GAATGGACGA ATCAACTCCA	18180
TAAATGCCCG CATGAGTCAA TTTATCTTTT TGATTATTTT CACCATCTGC GATTGGTACA	18240
TAGGCATCAC GTGGCGTTGT GGTCAGAGG ATTTTCTTGG TATCTCGATT GACAGTCATC	18300
AGGATGTTGA CATCTGATCG CGACACCGAA CTAATAGGAC CATAGGTGTC AATTCCACTA	18360
ACATAGATAT TGAAAGACTG ACTCTTAGAC GTCCTAGGAG CTTCTACTTT TTTAGTGAAT	18420
CCCTTAGTAT AAATCTTTTT TATCTTCGAT GCGTAGTCTG GATACTCTGA CTCGATGATG	18480
TTTTCAAAGA CACTATTTAG GACAATGGCC TTAGTCTCCC CTGCAATCAA ACTCTTGTA	18540
GCTGCCAAGT AAGACGAACT CTGGTTGACC GTCAAATCGG TATTCTGACT TGACTTGATA	18600
TCAGCTAGTA ATTTCTGAAT ATTTTCATTA TTAGTCCCAG TCGGTGCTGT CACACTCGTC	18660
AGTTGCGTAA CATTTTCGAT CTCACTATCT GCTAAAACAG CGACACTGAT TGAATATTCT	18720
GAGTAATTAG AAGTCGCATT TAAACGATTG GTCAGTCCAA CAAACTGCTG TACTGCAAAG	18780
AGCGACACAG AGCTGACAAG GATAGAGAAC ACCAACAGAA AAATAGTAAA CTTTTCAGCT	18840
TTTTTATAGA TAATCAAGAG TAGCCCTACC AAGGCAACTA GTAGGACTAA CGCAGTTACC	18900
ACTAGATTAA GATATCTAAA AGCAAGGATA TTGTACTTAA AGATTAAGAA CAATAAAAAA	18960
CAAATAACA ATAAATAAAT AGTCAGCAAA ACTATATTAA CACTTCGCTT CACTTTCTGT	19020
GAACGTGATT TTTTAAAACG TCTACTCATG ATTAATACCT ATACATTGAA CATTATACGA	19080
TTATATCACT TTTTACGGT AATGTCTACA CCTTTATTTT TACTATCTGC ATCTTTAAGT	19140
ATCTTAGTAG ACTTCCCGCG AAACAAAAAT ATAGTAAAAT GAAATAAGAA CAGAACAAAT	19200
CGTTCAGGAC AGTCAAATCG ATTTCTAACA ATGTTTGTAGA AGCAGAGGTG	19250

(2) INFORMATION FOR SEQ ID NO: 36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21706 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AAAGTTGAAA GACTGCTAGC TGTTTTTGAT ACCAATCGTT TCCAACTACA GAGCAAACAG	60
TATACAAAGT TTGTTTTTGG ATGTAAGCTT CTTGATGGAC AATTCCAAGA AAATCAAGAA	120
ATTGCTGACC TTCAATTTTT TGCCATTGAC CAACTGCCGA ACTTATCTGA AAAACGCATT	180
ACCAAGGAGC AAATAGAGCT TCTTTGGCAG GTTTATCAAG GTCATAGGGG GCAATATCTT	240

GACTAAGAAG ATGATTATCG TATTTCTAAA TCCATTTTTA ACAACTAGCA TGGTATAATA 300
ATATGCAGGA AAATTTTGAA TTATGAGGAA GACTAGATGA ATTTATGGGA TATTTTCTTT 360
ACGACTCAGG CAACCGAGCC GCCCAAATTT GACCTTTTTT GGTATGTTAG CCTATTTACG 420
CTCTTAGCCT TAACCTTTTA TACAGCCCAT CGCTATCGTG AAAAGAAGGT TTACCAACGA 480
TTTTTCCAAA TCTTGCAGAC TGTTCACTTA ATCCTTCTTT ATGGTTGGTA CTGGGTCAAT 540
CATATGCCAC TGTCAAGAAAG CCTACCTTTT TACCATTGCC GTATGGCTAT GTTGTGGTA 600
CTCTTGCTTC CTGGTCAATC CAAATATAAA CAATACTTTG CATTATTGGG AACATTTGGG 660
ACATTAGCAG CCTTTGTTTA TCCAGTGCCA GATGCTTACC CTTTCCACA TATCACCATT 720
CTATCCTTTA TCTTTGGTCA TTTAGCACTC TTGGGGAAC CTCTAGTTTA TCTATTGAGA 780
CAGTATAATG CGCGATTGCT GGATGTGAAG GGAATTTTC TCATGACCTT TGCCCTAAAT 840
GCCTTGATTT TTGTGGTCAA TTTGGTGACA GGTGGCGATT ACGGATTTTT GACAAAACCG 900
CCATTGGTTG GGGATCACGG TCTAGTAGCT AATTATTTAC TTGTTCAAT TGTGCTGGTA 960
GCTACTATCA GTTGGACTAA GAAAATCTTA GAATTCCTTT TAGCTCAAGA AGCAGAAAAA 1020
ATGATTGCAA AGGAAGCTTA ACACAGAGCT TTCTTTTTTG CTCTTAGAGA GTTTTACAA 1080
GCAGCTTATA AAATAAGAAT TTCTGAATAG ACAAACTCAA AAAATGGCTG GGAAATTTAG 1140
GAAAAAGCA AGCACGATTA AATTTTTTGT GTTATAATAT TTTGTGAATA GCTATGCCTA 1200
TGTTTAGCTA TGGAATAATA CGAAGTGCGA AACTTGGAAG ATAGAGAGGA AGCGATGTAA 1260
TGGCTAGAGA AGGCTTTTTT ACAGGTCTAG ATATTGGAAC AAGCTCTGTC AAGGTGCTTG 1320
TGGCCGAGCA GAGAAATGGT GAATTAAATG TAATTGGCGT GAGTAATGCC AAAAGTAAAG 1380
GTGTAAAGGA TGGAATTATT GTTGATATTG ATGCAGCAGC AACTGCTATC AAGTCAGCCA 1440
TTTCCAAGC GGAAGAAAAG GCAGGCATTT CGATTAAATC AGTGAATGTC GGCTTGCTTG 1500
GTAATCTTTT GCAGGTAGAA CCAACTCAGG GGATGATTCC AGTAACATCT GATACTAAGG 1560
AAATTACGGA TCAAGATGTT GAAAATGTTG TCAAATCAGC TTTGACAAAG AGTATGACAC 1620
CTGACCGTGA AGTCATTACC TTTATCTCTG AAGAATTTAT TGTGGATGGT TTCCAAGGGA 1680
TTCGTGACCC ACGTGGCATG ATGGGGGTC GCCTTGAAAT GCGTGGTTTG CTTTATACAG 1740
GACCTCGTAC TATCTTGCAC AATTTCGTA AGACGGTTGA GCGTGCAGGT GTTCAGGTTG 1800
AAAATGTTAT CATTTACCA CTAGCAATGG TTCAGTCTGT TTTGAACGAA GGGGAACGTG 1860
AATTGTTGTC TACAGTGATT GATATGGGG CAGGTCAAAC GACTGTGCT ACAATCCGTA 1920
ATCAAGAACT CCAGTTCACA CATATTCTCC AAGAAGGTGG AGATTATGTA ACTAAAGATA 1980

360
TCTCCAAGGT TTTGAAAACC TCTCGCAAAT TAGCGGAAGG CTTGAAACTG AATTACGGGG 2040
AAGCCTATCC GCCTCTTGCA AGCAAAGAAA CCTTCCAAGT AGAGGTTATT GGAGAAGTAG 2100
AAGCAGTCGA AGTGACGGAA GCCTACTTGT CAGAAATTAT TTCTGCACGA ATCAAGCACA 2160
TCCTTGAACA AATCAAGCAA GAATTAGATA GAAGGCGTCT ATTGGACCTC CCTGGTGGTA 2220
TTGTCTTAAT CGGTGGGAAT GCCATTTTAC CAGGTATGGT TGAGCTTGCT CAGGAAGTCT 2280
TTGGCGTCCG TGTCAAGCTT TATGTTCCAA ATCAAGTTGG TATCCGTAAT CCAGCCTTTG 2340
CGCATGTGAT TAGTTTATCA GAATTTGCGG GTCAATTAAC AGAAGTTAAT CTTTGGCTC 2400
AGGGAGCGAT AAAAGGTGAG AATGACTTAA GTCATCAGCC AATTAGTTTT GGTGGGATGC 2460
TGCAAAAAAC AGCTCAGTTT GTACAATCAA CGCCTGTTCA ACCAGCTCCT GCTCCAGAAG 2520
TAGAGCCGGT GGCGCCTACA GAACCAATGG CGGATTTCCA ACAAGCTTCA CAAAATAAAC 2580
CGAAATTAGC AGATCGTTTC CGTGGATTGA TCGGAAGCAT GTTTGACGAA TAAAGAGGAA 2640
AAATAAATTA TGACATTTTC ATTTGATACA GCTGCTGCTC AAGGGGCAGT GATTAAAGTA 2700
ATTGGTGTGC GTGGAGGTGG TGGCAATGCC ATCAACCGTA TGGTCGACGA AGGTGTTACA 2760
GGCGTAGAAT TTATCGCAGC AAACACAGAT GTACAAGCAT TGAGTAGTAC AAAAGCTGAG 2820
ACTGTTATTC AGTTGGGACC TAAATTGACT CGTGGTTTGG GTGCAGGAGG TCAACCTGAG 2880
GTTGGTCTGA AAGCCGCTGA AGAAAGCGAA GAAACACTGA CGGAAGCTAT TAGTGGTGCC 2940
GATATGGTCT TCATCACTGC TGGTATGGGA GGAGGCTCTG GAACTGGAGC TGCTCCTGTT 3000
ATTGCTCGTA TCGCCAAAGA TTTAGGTGCG CTTACAGTTG GTGTTGTAAC ACGTCCCTTT 3060
GGTTTTGAAG GAAGTAAGCG TGGACAATTT GCTGTAGAAG GAATCAATCA ACTTCGTGAG 3120
CATGTAGACA CTCTATTGAT TATCTCAAAC AACAATTTGC TTGAAATTGT TGATAAGAAA 3180
ACACCGCTTT TGGAGGCTCT TAGCGAAGCG GATAACGTTT TCGTCAAGG TGTTCAGGG 3240
ATTACCGATT TGATTACCAA TCCAGGATTG ATTAACCTTG ACTTTGCCGA TGTGAAAACG 3300
GTAATGGCAA ACAAAGGGAA TGCTCTTATG GGTATTGGTA TCGGTAGTGG AGAAGAACC'T 3360
GTGGTAGAAG CGGCACGTAA GGCAATCTAT TCACCACTTC TTGAAACAAC TATTGACGGT 3420
GCTGAGGATG TTATCGTCAA CGTTACTGGT GGTCTTGACT TAACCTTGAT TGAGGCAGAA 3480
GAGGCTTCAC AAATTGTGAA CCAGGCAGCA GGTCAAGGAG TGAACATCTG GCTCGGTACT 3540
TCAATTGATG AAAGTATGCG TGATGAAATT CGTGTAACAG TTGTTGCAAC GGGTGTTCGT 3600
CAAGACCGCG TAGAAAAGGT TGTGGCTCCA CAAGCTAGAT CTGCTACTAA CTACCGTGAG 3660
ACAGTGAAAC CAGCTCATTC ACATGGCTTT GATCGTCATT TTGATATGGC AGAAACAGTT 3720
GAATTGCCAA AACAAAATCC ACGTCGTTTG GAACCAACTC AGGCATCTGC TTTTGGTGAT 3780

361

TGGGATCTTC GCCGTGAATC GATTGTTTCGT ACAACAGATT CAGTCGTTTC TCCAGTCGAG	3840
CGCTTTGAAG CCCCAATTTC ACAAGATGAA GATGAATTGG ATACACCTCC ATTTTTCAAA	3900
AATCGTTAAG TAAATGAATG TAAAAGAAAA TACAGAACTT GTTTTTCGAG AAGTTGCAGA	3960
GGCTAGTCTG AGTGCTCATC GAGAGAGTGG TTCGGTCTCT GTCATTGCAG TTACCAAGTA	4020
TGTAGATGTA CCGACAGCGG AAGCCTTGCT TCCGCTAGGT GTCCATCATA TCGGTGAAAA	4080
TCGTGTAGAT AAGTTTCTGG AAAAATATGA AGCTTTAAAA GATCGAGATG TGACTTGGCA	4140
TTTGATTGGT ACCTTGCAAA GACGTAAGGT GAAAGATGTC ATTCAATACG TTGATTATTT	4200
CCATGCATTG GACTCAGTAA AGCTAGCAGG GGAAATTCAA AAAAGAAGTG ACCGAGTCAT	4260
CAAGTGTTTC CTTCAAGTAA ATATTTCTAA AGAAGAAAGC AAACACGGTT TTCGAGAGA	4320
GGAAGTCTG GAAATCTTGC CAGAGTTAGC CAGACTAGAT AAGATTGAAT ATGTTGGTTT	4380
AATGACGATG GCACCTTTTG AGGCTAGCAG TGAGCAGTTG AAAGAGATTT TCAAGCGGC	4440
CCAAGATTTA CAAAGAGAAA TTCAAGAGAA ACAAATCCA AATATGCCTA TGACCGAGTT	4500
AAGTATGGGA ATGAGTCGTG ATTATAAAGA AGCGATTCAA TTCGGTTCCA CTTTGTTCG	4560
TATAGGTACA TCATTTTTTA AGTAGGAGAG AACCATGTCT TTAAGAGATA GATTGATAG	4620
ATTTATAGAT TATTTTACGG AGGATGAGGA TTCAAGTCTC CCTTATGAAA AAAGAGATGA	4680
GCCTGTGTTT ACTTCAGTAA ATTCTTCACA GGAACCGGCT CTCCAATGA ATCAACCTTC	4740
ACAGTCGGCT GGCACAAAAG AGAACAATAT CACCAGACTT CATGCAAGAC AACAGGAATT	4800
GGCAAATCAG AGTCAGCGTG CAACGATAA GGTCATTATA GATGTTTCGT ATCCTAGAAA	4860
ATATGAGGAT GCAACAGAAA TTGTTGATT ATTGGCAGGA AACGAAAGTA TCTTGATTGA	4920
TTTTCAGTAT ATGACAGAGG TGCAGGCTCG TCGTTGTTTG GACTATTTGG ATGGAGCTTG	4980
TCATGTTTTA GCTGGAAATT TGAAAAAGGT AGCTTCTACC ATGTATTTGT TGACACCAGT	5040
GAACGTTATT GTAAATGTTG AAGATATCCG TTTACCAGAT GAAGATCAAC AGGGTGAGTT	5100
CGGTTTTGAT ATGAAGCGAA ATAGAGTACG ATAATGATTT TTTTAATTTC TATGATTTAT	5160
AATGCAGTGG ATATTTACTC CCGATTTTG GTAGCCTTCG CTGTCATGTC TTGGTTTCCA	5220
GGTGCCTACG AATCCAGTTT AGGTCGTTGG ATTGTAGCGT TGGTGAAACC AGTGCTTGCT	5280
CCCTTGCAAC GCCTGCCTTT ACAGATAGCG GGTCTTGATT TATCTGTTTG GGTGCGATT	5340
GTTTTGGTTC GATTTTTAGG AGAAAACCTA GTGCGTTTTTC TGGCGATGAT AGGATGAATA	5400
AAGGGATTTA TCAGCATTTT TCCATAGAAG ATCGTCCATT TCTTGACAAG GGAATGGAAT	5460
GGATAAAGAA GGTAGAAGAT AGCTATGCTC CTTTTTTAAC TCCTTTTATC AATCCTCATC	5520

362

AGGAGAAGCT ATTAAAGATT TTGGCCAAAA CCTATGGTCT TGCTTGTAGC AGTAGTGGG	5580
AATTCGTCTC GAGTGAGTAT GTTCGAGTTT TATTATACCC AGATTATTTT CAACCAGAGT	5640
TTTCAGATTT TGAAATATCT CTCCAGGAAA TTGTGTATTC CAATAAATTT GAACATTTAA	5700
CGCATGCTAA GATTTTAGGG ACAGTCATCA ATCAATTAGG GATTGAACGG AAACTTTTTG	5760
GAGATATCCT AGTAGATGAA GAACGGGCGC AGATTATGAT TAATCAGCAG TTTCTTCTTC	5820
TCTTTCAAGA TGGACTAAAG AAAATTGGTC GTATACCTGT TTCGCTGGAG GAACGTCCTT	5880
TCACCGAGAA AATAGATAAG CTAGAACAGT ATCGAGAACT GGATTTATCT GTGTCTAGTT	5940
TTCGATTAGA TGTCTTTTA TCAAAATGTT TGAAACTATC TAGGAATCAA GCAAACCAGT	6000
TGATTGAAAA GAAACTTGTC CAAGTAAAT ATCATGTGGT AGACAAATCA GATTACACTG	6060
TTCAAGTTGG AGACTTGATT AGTGTGAGAA AATTTGGTCG CTTGAGATTA CTTCAAGATA	6120
AGGGACAAAC GAAAAAGAG AAGAAAAAA TAACCGTCCA GTTATTATTA AGTAAGTGAG	6180
GAATAGAATG CCAATTACAT CATTAGAAAT AAAGGACAAG ACTTTTGAA CTCGATTCAG	6240
AGGTTTIGAT CCAGAAGAAG TCGATGAATT TTAGATATT GTGGTTCGTG ATTACGAAGA	6300
TCTTGTGCGT GCGAATCATG ATAAAAATTT GCGTATTAAG AGTTTAGAAG AGCGTTTGTC	6360
TTACTTTGAT GAAATAAAAG ATTCATTGAG CCAGTCTGTA TTGATTGCTC AGGATACAGC	6420
TGAGAGAGTG AACAGGCGG CGCATGAACG TTCAAACAAT ATCATTTCATC AAGCAGAGCA	6480
AGATGCGCAA CGCTTGTGG AAGAAGCTAA ATATAAGGCA AACGAGATTC TTCGTCAAGC	6540
AACTGATAAT GCTAAGAAAG TCGCTGTTGA AACAGAAGAA TTGAAGAACA AGAGCCGTGT	6600
CTTCCACCAA CGTCTCAAAT CTACAATTGA GAGTCAGTTG GCTATTGTTG AATCTTCAGA	6660
TTGGGAAGAT ATTCTCCGTC CAACAGCTAC TTATCTTCAA ACCAGTGATG AAGCCTTTAA	6720
AGAAGTGGTT AGCGAAGTAC TTGGAGAACC GATTCCAGCT CCAATTGAAG AAGAACCAAT	6780
TGATATGACA CGTCAGTTCT CTCAAGCAGA AATGGCAGAA TTACAAGCTC GTATTGAGGT	6840
AGCCGATAAA GAATTGCTCG AATTGGAAGC TCAGATTAAA CAGGAAGTGG AAGCTCCAAC	6900
TCCTGTAGTG AGTCCTCAAG TTGAAGAAGA GCCTCTGCTC ATCCAGTTGG CCCAATGTAT	6960
GAAGAACCAG AAGTAGCTCC AATGCATCCG ATAGGTCCAA CACCAGCTAC AGAACTGTT	7020
GATTCAATAC CGGATTTGA AGCACCGCAA GAATCTGTTA CAATTTTATA AGAAATATTC	7080
TGAGAACAAT ATCTTATCCT TATATTTCCA GCGAGCAGGA GATGGTGTGA GTCCTGTAAT	7140
CCCTATTGAT AAGATTATCC TCTCAAAAAC TCAAGTCTGA AGCTAGTAAG ATTTGACGTT	7200
TCCCACGTTA CGGATAAGA GGGAGAAAGA CTAAATCTTT TTCCGAATAA AGGTGGTACC	7260
ACGATTTTCG TCCTTTTGG AAGTCGTGGT TTTTAATTTG TTATTATTTA TAAAGGAGAT	7320

363

ACCATGAAAC TCAAAGACAC CCTTAATCTT GGGAAAACTG AATTCCCAAT GCGTGCAGGC	7380
CTTCCTACCA AAGAGCCAGT TTGGCAAAAG GAATGGGAAG ATGCAAAACT TTATCAACGT	7440
CGTCAAGAAAT TGAACCAAGG AAAACCTCAT TTCACCTTGC ATGATGGCCC TCCATACGCT	7500
AACGGAAATA TCCACGTTGG ACATGCTATG AACAAGATTT CAAAAGATAT CATTGTTTCGT	7560
TCTAAGTCTA TGTCAAGATT TTACGCACCA TTTATTCTTG GTTGGGATAC TCATGGTCTG	7620
CCAATCGAGC AAGTCTTGTC AAAACAAGGT GTCAAACGTA AAGAAATGGA CTTGGTTGAG	7680
TACTTGAAAC TTGCGGTGA GTACGCTCTT TCTCAAGTAG ATAAACAACG TGAAGATTTT	7740
AAACGTTTGG GTGTTTCTGG TGAATGGGA AATCCATATG TGACCTTGAC TCCTGACTAT	7800
GAAGCAGCTC AAATTCGTGT ATTTGGTGAG ATGGCTAATA AGGGTTATAT CTACCGTGGT	7860
GCTAAGCCAG TTTACTGGTC ATGGTCATCT GAGTCAGCAC TTGCTGAAGC AGAGATTGAA	7920
TACCATGACT TGGTTTCAAC TTCCCTTTAC TATGCCAACA AGGTAAAAGA TGGCAAAGGA	7980
GTTCTAGATA CAGATACTTA TATCGTTGTC TGGACAACGA CTCCATTTAC CATCAGAGCT	8040
TCTCGTGGTT TGACGGTTGG TGCAGATATT GATTACGTTT TGGTTCAACC TGCTGGTGAA	8100
GCTCGTAAGT TTGTCGTTGC TGCTGAATTA TTGACTAGCT TGTCTGAGAA ATTTGGCTGG	8160
GCTGATGTTT AAGTTTTGGA AACTTACCGT GGCCAAGAAC TCAACCACAT CGTAACAGAA	8220
CACCCATGGG ATACAGCTGT AGAAGAGTTG GTAATTCTTG GTGACCACGT TACGACTGAC	8280
TCTGGTACAG GTATTGTCCA TACAGCCCCT GGTTTTGGTG AGGACGATTA CAATGTTGGT	8340
ATTGCTAATA ATCTTGAAGT CGCAGTGAAT GTTGATGAAC GTGGTATCAT GATGAAGAAT	8400
GCTGGTCCTG AATTTGAAGG TCAATTCTAT GAAAAGGTAG TTCCAACGTG TATTGAAAAA	8460
CTTGGTAACC TCCTTCTTGC CCAAGAAGAA ATCTCTCACT CATATCCATT TGAATGGCGT	8520
ACTAAGAAAC CAATCATCTG GCGTGCAGTT CCACAATGGT TTGCCTCAGT TTCTAAATTC	8580
CGTCAAGAAA TCTTGGACGA AATTGAAAAA GTGAAATTC ACTCAGAATG GGGTAAAGTC	8640
CGTCTTTACA ATATGATCCG TGACCGTGGT GACTGGGTTA TCTCTCGTCA ACGTGCTTGG	8700
GGTGTTCAC TTCTATCTT CTACGCTGAA GATGGTACAG CTATCATGGT AGCTGAAACT	8760
ATTGAACACG TAGCTCAACT TTTTGAAGAA TATGGTTCAA GCATTTGGTG GGAACGTGAT	8820
GCCAAAGACC TCTTGCCAGA AGGATTTACT CATCCAGGTT CACCAAACGG CGAGTTCAAA	8880
AAAGAAACTG ATATCATGGA CGTTTGGTTT GACTCAGGTT CATCATGGA TGGAGTGGTG	8940
GTAAACCGTC CTGAATTGAC TTACCCAGCC GACCTTTACC TAGAAGGTTG TGACCAATAC	9000
CGTGGTTGGT TTAACATCAT ACTTATCACA TCTGTTGCCA ACCATGGCGT AGCACCTTAC	9060

364

AAACAAATCT TGTCACAAGG TTTTGCCCTT GATGGTAAAG GTGAGAAGAT GTCTAAATCT	9120
CTTGGAAATA CTATTGCTCC AAGCGATGTT GAAAAACAAT TCGGTGCTGA AATCTTGCGT	9180
CTCTGGGTAA CAAGTGTGA CTCAAGCAAT GACGTGCGTA TCTCTATGGA TATCTTGAGC	9240
CAAGTTTCTG AAACCTACCG TAAGATTCGT AACACTCTTC GTTCTTGAT TGCCAATACA	9300
TCTGACTTTA ACCCAGCTCA AGATACAGTC GCTTACGATG AGCTTCGTTC AGTTGATAAG	9360
TACATGACGA TTCGCTTTAA CCAGCTTGTC AAGACCATTG GTGATGCCTA TGCAGACTTT	9420
GAATTCTTGA CGATCTACAA GGCCTTGGTG AACTTTATCA ACCTTGACTT GTCAGCCTTC	9480
TACCTTGATT TTGCCAAAGA TGTGTTTAC ATTGAAGGTG CCAAATCACT GGAACGCCGT	9540
CAAATGCAGA CTGTCTTCTA TGACATTCTT GTCAAAATCA CCAAACTCTT GACACCAATC	9600
CTTCCTCACA CTGCGGAAGA AATCTGGTCA TATCTTGAGT TTGAAACAGA AGACTTCGTC	9660
CAATTGTCAG AATTACCAGA AGTTCAAACT TTTGCTAACC AAGAAGAAAT CTGGGATACA	9720
TGGGCAGCCT TCATGGACTT TCGTGGACAA GCACAAAAG CCTTGAAGA AGCTCGTAAT	9780
GCAAAAGTTA TCGGTAAATC ACTTGAAGCA CACTTGACAG TTTATCCAAA TGAAGTTGTG	9840
AAACTCTAC TCGAAGCAGT AAACAGCAAT GTAGCACAAC TTTTGATCGT GTCTGAGTTG	9900
ACCATCGCAG AAGGACCAGC TCCGGAAGCT GCCCTTAGCT TCGAAGATGT AGCCTTCACA	9960
GTGAACGTG CTACTGGTGA AGTATGTGAC CGTTGCCGTC GTATCGACCC AACAACAGCA	10020
GAACGCAGCT ACCAGGCAGT TATCTGTGAC CACTGTGCAA GCATCGTAGA AGAAAACTTT	10080
GCGGAAGCAG TCGCAGAAGG ATTTGAAGAG AAATAAGATT GAAAAGTCTA GGCAAAATTC	10140
AATTTGAGAA GAAAAGACAA CTAATTTTAT AGTCTATTAA ACGCATTGTA TCACGTTTTT	10200
GAATACCTGA TATGATGCGT TTTTATTTTA TTTTAAAAAT TTGCGAGGTA TGACTTTTTA	10260
TACTCAACAA GAATCAAAGA GAAACTTAGC AAGCTAACAG TAGTAAGATA AAATAGGAAT	10320
TTGATATTAG GGATAAGATT GGTAAATAGT GTAATTTTTT TACAACAATA AATTTATATA	10380
GTTATTTCTG GTTCTGAAA AGTATTATAT TTTATTTCAT ATTATACAAA TTTTATTTTT	10440
ATAATATCAG AACATACTTT TTTTAAAAGC AAATATGATA CAATTTTATT TGAAAAAAT	10500
AAAAAAGGAG ATTTTATTAT AAAATTAAAA AGACTTGCTT TAATTAGTGG TATCGTCGGT	10560
CTTGTGGGAG GAATTTTACT TCTTATTGGT CCTTTTGTCT TGTGGGAAT AGCGGTAAAC	10620
ACAGCTGCTA CAACTCTTAA TGGAGGAGCT ACTGCAGGGG CTTTTCAGG TGAGCCTTA	10680
CTCTTGAATG CCTTGAAGAT TGCAAATCTT GTTCTTGGA TCATTGCTAT TGTTTACTAT	10740
AAAGGAGATA AGCGTGTAGG TGCAGCTCCG TCTGTACTAA TGATTGTTTC TGGTGGAGTT	10800
AGTCTCATTC TATCCGTTT TTAGGATGGG TTGGGGGGAT TTTTGCTATT ATCGGAGGAT	10860

365

CTCTATTCCT TTCAACATTG AAGAAATTCA AATCAGAAGA ATAAAAGGTA TTTTAGCATG	10920
AAAAGAACAA AAAAGTTTAT CGGTATAGGA GTAGCTCTAT TATCTCTTTC TCTTCTAGTT	10980
GCATGTGGAA CATAAAGTTC AAAGAATACT TCAACAAGTA ATGATGAGAA GACAGTAGCA	11040
ACATCCAATA GTTCAAAAGA AACAACTACT TTCGATACAC CGGTTGTAAC AGACGATGCG	11100
ATTGAATCAA TACGCACTTA TGCAGATTAT ATAGATCTTT ATAAAAATAT TTTTGATGAT	11160
TATTTTACTA AAGCTGAGGA AGGTTTCAAA GGCATAGCTA TGGAAAATAA TGACTCGTTT	11220
ACTAAACTAA AAGAGTCAAC TCAAAAATTA TTCGATGCGC AGAAAAAAG GTTAAATAAT	11280
GAAGATAGAA TAGAAACAAC CAAAAACAAT GTGATTGCCA AACATTGTCA AACAGTCCTT	11340
TCCTTTTGG TMTTGACTAG CTTTTTTGTG AAAAAATTGTG TAAAAAGAA TAGATAAACG	11400
AGGGGAAACC TCGGAAAATT TAAAGGAGAA TCCATCTAAT GGTAAAATTG GTTTTTGCTC	11460
GCCACGGTGA GTCTGAATGG AACAAAGCTA ACCTTTTCAC TGGTTGGGCT GATGTTGATT	11520
TGTCTGAAAA AGGTACACAA CAAGCGATTG ACGCTGGTAA ATGATCAAA GAAGCTGGTA	11580
TCGAATTTGA CCAAGCTTAC ACTTCAGTAT TGAAACGTGC TATCAAAACA ACTAAGTTGG	11640
CTCTTGAAGC TTCTGACCAA TTGTGGGTTT CAGTTGAAAA ATCATGGCGC TTGAACGAAC	11700
GTCCTACGG TGGTTTGACT GGTA AAAACA AAGCTGAAGC TGCTGAACAA TTTGGTGATG	11760
AGCAAGTTCA CATCTGGCGT CGTTCATACG ATGTATTGCC TCCAAACATG GACCGTGATG	11820
ATGAGCACTC AGCTCACACA GACCGTCGTT ACGCTTCACT TGACGACTCA GTTATCCCAG	11880
ATGCTGAAAA CTTGAAAGTG ACTTTGGAAC GTGCTCTTCC ATTCTGGGAA GATAAAATCG	11940
CTCCAGCTCT TAAAGATGGT AAAAACGTAT TCGTAGGAGC TCACGGTAAC TCAATCCGTG	12000
CCCTTGTA AAACATCAAA GGTTCGTCAG ATGACGAGAT CATGGACGTG GAAATCCCTA	12060
ACTTCCCACC ATTGGTATTC GAATTCGACG AAAAAATTGAA CGTCGTTTCT GAATACTACC	12120
TTGGAAAATA AAAAAATTGTA AGTCTAGAAT TGATTTCTAG GCTTTTATG TTAGTATGGA	12180
AGTATGATAA GGAATAAAAA ACAAGATTAT GTACTGGCCT ACAAGCAACC AGCTTCAACC	12240
ACTTACATGG GTTGGGAAGA AGAAGCTTTA CCGATAGGCA ATGGTTCTTT AGGAGCAAAA	12300
GTATTTGGCC TTATAGGGGC TGAACGGATT CAATTTAATG AAAAAAGTCT CTGGTCTGGA	12360
GGTCCACTTC CTGATAGTTC AGATTATCAG GGTGGAAATC TTCAGGATCA GTATGTTTTT	12420
TTAGCTGAGA TTCGGCAGGC TTTGGAGAAG AGAGATTACA ATCTGGCTAA GGAAGTGGCT	12480
GAGCAGCACC TAATTGGGCC AAAACGAGT CAATATGGGA CCTATCTGTC TTTTGGGGAT	12540
ATTCACATTG AGTTCAGCCA GCAAGGTACG ACTTTGTCTC AGGTGACGGA CTATCAGAGA	12600

366

CAGCTGAATA TTAGTAAGGC ACTTGCACG ACITCTTATG TCTATAAGGG AACGCGATTT	12660
GAACGTAAAG CTTTTCGAG TTTTCCAGAT GATCTCTTGG TTCAATGTTT TACTAAGGAA	12720
GGGTTGGAAA CTCTAGATTT TACTATAGAA CTATCCTTGA CCTGTGATTT GGCTTCTGAT	12780
GGAAAGTATG AGCAGGAAAA ATCTGATTAC AAGGAGTGTA AGTTGGATAT TACTGATTCT	12840
CATATCTTGA TGAAGGGAAG AGTTAAGGAT AATGATCTGC GGTTCGCTAG TTATCTAGCT	12900
TGGGAAACGG ATGGAGATAT TAGAGTTTGG TCAGATAGGG TTCAGATATC AGGAGCCAGT	12960
TATGCCAATC TCTTCTTGGC CGCTAAGACG GATTTTGCCC AAAATCCTGC TAGCAATTAT	13020
CGCAAGAAAC TAGATTTAGA GCAACAGGTG ATAGACTTGG TGGACACAGC TAAAGAAAAG	13080
GGCTATACCC AATTGAAATC AAGGCATATC GAGGACTACC AAGCCTTATT CCAGCGTGT	13140
CAATTGGATT TGGAAGCTGA TGTGACGCA TCCACTACAG ATGATTTGTT AAAAAATTAT	13200
AAGCCACAAG AAGGCGAGGC TTTGGAGGAG CTGTTCTTCC AGTATGGACG GTATTTATTG	13260
ATTAGTTCGT CCAGAGACTG CCCAGATGCT CTACCAGCTA ACCTACAGGG AGTCTGGAAT	13320
GCGGTCGACA ATCCTCCTTG GAATTCGGAC TATCACTTAA ATGTCAATCT GCAGCTGAAT	13380
TATTGGCCAG CCTATGTTAC CAATCTCCTA GAGACGGTCT TTCCAGTCAT CAACTATGTA	13440
GATGATTTGC GTGTCTATGG TCGTCTAGCG GCTGTAAAGT ATGCAGGAAT CGTCTCTCAG	13500
AAAGGTGAGG AGAATGGTTG GTTGGTTCAT ACTCAAGCGA CTCCTTTTGG TTGGACGGCA	13560
CCTGGTTGGG ATTACTATTG GGGTTGGTCA CCAGCTGCCA ATGCGTGGAT GATGCAAAAC	13620
GTTTATGAAG CCTATTTATT TTATAGGGAC CAAGACTATC TCAGGGAGAA AATTTATCCC	13680
ATGTTGAGGG AACCGTTCG TTTTGGGAAT GCCTTTTAC ATAAGGATCA GCAGGCGCAG	13740
CGTTGGGTGT CTTCTCCGTC TTATTCCCCA GAACATGGGC CGATTTTCGAT TGGCAATACC	13800
TATGACCAAT CTCTGATTTG GCAGTTATTT CATGATTTTA TTCAGGCTGC TCAGGAATTG	13860
GGACTGGATG AGGACTTGTT GACTGAGGTT AAGGAGAAGT CTGATTTACT AAATCCTTTG	13920
CAAATCACTC AATCTGGTCG AATCAGGGAG TGGTATGAGG AGGAAGAGCA GTATTTTCAA	13980
AATGAGAAAG TGGAGGCCCA GCATCGGCAC GCTTCCCATC TAGTGGGACT CTATCCTGGC	14040
AATCTCTTTA GCTACAAGGG ACAAGAGTAT ATTGAAGCGG CGCGTGCTAG CCTCAATGAT	14100
CGTGGAGATG GCGGCACAGG CTGGTCCAAG GCTAATAAGA TCAATCTCTG GCGCGTTTG	14160
GGAGATGGCA ATCGAGCCCA TAAATTATTG GCAGAGCAGT TAAAGACATC CACCTTGCAA	14220
AATCTTTGGT GTAGCCATCC TCCTTTTCAG ATAGATGGTA ATTTTGGTGC TACTAGTGGC	14280
ATGGCAGAAA TGTTACTCCA GTCTCATGCA GCTTATCTGG TACCTCTAGC TGCCCTACCT	14340
GATGCTTGGT CAACAGGTTT TGTTCAGGC TTAATGGCAC GTGGACATTT TGAAGTGAGC	14400

367

ATGAGCTGGG AAGATAAAAA ACTCTTACAG TTGACCATTT TATCAAGGAG TGGAGGAGAT	14460
TTGCGAGTTT CTTATCCAGA TATTGAGAAG AGTGTGATTA AAATGAATCA AGAAAAATA	14520
AAAGCGAAAT GCATGGGGAA AGATTGTATT TCGGTGGCAA CAGCAGAAGG TGATCTTGTT	14580
CAATTTTATT TTTAAGAAGA TGTATAAGG CAGTAATTTG AACTGCCTT TTAATAAGGA	14640
TTTAAGAATA TAAGCAGTTT TCAACTAGTT GAAAAACGT TATAATGATA ATAGGAAGTA	14700
ATACTCAATG AAAATCAAAG AGCACAACT AGGAAGCTAG CCGCAGGTG CTCAAAACAG	14760
TGTTTTGAGG TTGCAGATGG AAGCTGACGT GGTTTGAAGA GAGATTTTCG AGGAGTATAA	14820
TTTGTTTGAT AGAGGGTGGG TCTGATGGCT TATATTGAGA TGAAACACTG TTACAAGCGT	14880
TATCAGGTG GGGACACGGA GATTGTGGCC AATTGTGATG TGAATTTTGA GATTGAAAAG	14940
GGGGAGCTGG TTATTATCCT TGGTGCTTCA GGTGCAGGCA AGTCAACAGT TCTTAACCTT	15000
CTTGGGGGAA TGGATACCAA TGATGAAGGG GAAATCTGGA TTGATGGTGT TAATATTGCG	15060
GATTATAGTT CCCACCAGCG CACCAATTAC CGTAGAAATG ATGTGGGGTT TGTTTTTCAG	15120
TTTTATAATC TAGTTTCTAA TCTGACAGCT AAGGAAAATG TGGAACGGC TTCTGAAATT	15180
GTGACAGATG CCTTGAATCC TGATCAGGCC TTGACAGATG TAGGTCTGGC TCATCGTCTC	15240
AATAACTTTC CAGCCCAGCT TTCTGGAGGG GAGCAACAGC GAGTCTCCAT TGCACGCGCG	15300
GTAGCCAAAA ATCCTAAAT TCTCCTTTGT GATGAACCGA CTGGAGCCTT GGATTATCAG	15360
ACGGGCAAGC AGGTTTGA AATTCTCCAA GACATGTCTC GTCAAAAGGG AGCGACGGTG	15420
ATCATCGTGA CTCATAATGG AGCTTTGGCG CCCATTGCTG ATCGCGTGAT TCAAATGCAC	15480
GATGCCAGTG TCAAGGATGT GGTGCTCAAC CAGCATCCTC AGGATATTGA CAGTTTGGAG	15540
TACTAGCATG ATCAAGCGAA AACTTATTG GAAGGACTTA GTTCAGTCCT TCACAGGCTC	15600
CAAGGGCGT TTTTATCCA TCTTGATCCT GATGATGTTG GGATCTCTAG CCTTAGTAGG	15660
CCTCAAAGTA ACCAGTCCCA ACATGGAGGC GACAGCTAAT GCTTATTTAA CAACTGCTCA	15720
AACCTTGGAT TTGGCAGTCA TGTCTAACTA TGGCTTGGAT CAAGCAGACC AAGAAGAACT	15780
AAAACAGACG GAGGGCGCAG AGGTCGAGTT TGGCTATTTG ACAGATGTGA CTATGGATAA	15840
TGGGCAGGAT GCCATTCGGC TGTACTCCAA ACCAGAGCGA ATTTCAACCT TTCAGCTAAG	15900
AAAGGGACGA CTTCTCAGT CAGACAAGGA AATCGCTTTG GCCACTCATT TGCAAGGCCA	15960
ATACAGCGTG GGACAGGAGA TTAGTTTAA AGAAAAAGAA GAGGGTCATT CCTCTTAAA	16020
AGACCATACT TATACCATTA CTGGTTTGT GGATTCGGCT GAAATCCTCT CCCAGCGAGA	16080
TATGGGCTAC GCAGGAAGTG GAAGTGGGAC TCTGACAGCC TATGGGGTGA TTTTACCTAG	16140

368

TCAATTTGAT CAGAAAGTCT ACAATATAGC TCGTTTGAAA TATCAAGATT TAGCGGGTTT	16200
AAATGCCTTT TCATCAGCTT ATGAAGAAAA ATCCAAGCAA CATCAAGAAG AGCTTGAACA	16260
AATTTTATCA GATAATGGCA AGGTACGTCT GCAACTTTTG AAAAAAGAAG GACAAGAGTC	16320
TCTAGACAAG GGGCAAGAGA CCCTTGACAA GGCTCAGACT AATTTCAGG AAGGCAAGCG	16380
TCGTTTAGCA GCTGCTCAAG CTCGTATACA GGCTCAAGAA AGTCAACTAG CCTTGTTC	16440
TCAAGTTCAG AGAGAGCAGG CTAGTGCTCA ACTTACCCAA GCCAAGCAGG AATTGGGCAA	16500
GGAAGAGGAC AACTAAAGC AAGCTGAACA AAATCTAGCC CAAGAAAAGG AAAAATTAGA	16560
AAAACATCAG CAAGTCTTGG ATGATTTGGC GGAGCCAAGG TATCAGGTTT ATAATCGTCA	16620
GACCATGCCA GGTGGTCAGG GCTATCTTAT GTATAGCAAT GCTTCATCCA GTATTCGAGC	16680
AGTGGGCAAT ATCTTTCCTG TGGTACTTTA TGCCGTAGCA GCCATGGTGA CCTTTACGAC	16740
CATGACTCGC TTTGTAGACG AAGAGCGAAC TCATGCAGGG ATTTTAAAGG CCTTGGGTTA	16800
TCGTAGTAAG GATATTATCG CCAAGTTTCT CCTTTATGGA CTAGTAGCTG GGACTGTGG	16860
AACGGCTCTA GGTAGTATAC TTGGTCATTA TTTGCTAGCC AGTGTAATTT CAAGTGTCAT	16920
TACAAAAGGC ATGGTGGTGG GAGAACTCA GATTCAGTTC TATTGGACCT ATAGCTTACT	16980
AGCTTTTGTC TTGAGCTTGT TGGCGAGTGT GTTACCAGCC TATCTGGTGG CTGGAGGGA	17040
ACTTCATGAC GAAGCAGCCC AGCTTCTACT TCCTAAACCT CCTGTCAAAG GAGCTAAAAT	17100
CTTATTGGAG CGTATCGGTT TTATCTGGCG TCGTCTCAGT TTTACTCATA AGGTAACAGC	17160
CCGCAACATC TTTCGTTATA AGCAGAGAAT GTTGATGACA ATCTTTGGTG TGGCAGGTTT	17220
TGTAGCTCTG CTCMTTGAG GTTTGGGAAT CCAATCTTCT GTAGCAGGAG TTCCGTCTAA	17280
ACAGTTTCAA CAAATCCAAC AGTATCAGAT GCTTGTCTCT GAAAATCCTA GTGCGACCAA	17340
TCAGGACAAG GTAGAGCTAG CAGAAGTGT GAAAGGGCAG GAGATACTAG CCTACCAGAA	17400
AATCTATTCT AAAGCGCTAT ACAAGGATTT CAAAGGCAAA GCTGGTCTTC AAAACATTAC	17460
TCTTATGATG ATAGAGAAGG AAGATTTGAC TCCCTTTATC CATCTTCAAC ATCATCAGCA	17520
GGAGCTGACA TTTAAAGATG GCATCGTTAT TACAGCTAAA CTCGCCCAGC TGGCAGGTGT	17580
CAAGGTTGGG CAGACTTTAG AAATTGAAGG TAAGGAACTA AAGGTCGTTG CTATTACTGA	17640
GAACTACGTT GGTCACTTTA TTTATATGAG TCAGGCTAGC TATGAGCAAC TTTACGGACA	17700
GCTACCCCAA GCCAACACTT ATCTGGTCTC ATTAAGGGAT ACCAGTGCAA CTAGTATCGA	17760
AAGTCAGGCG GGCTTGCTTA TGAATCAATC TCGGTTGTCC AGCGTTGTCC AAAATGCTTC	17820
AGCCATTGCA CTCTCGACT CTATCGCTAG CTCACTCAAT CAGACCATGA CCATCTGGT	17880
CATCGTATCG GTTCTATTAG CTATGTGTCAT CCTTTACAAT CTGACCAATA TCAACGTAGC	17940

TGAGAGAATC CGTGAACCTCT CCACTATCAA GGTTCTTGGT TTTCATAATA ATGAAGTCAC	18000
CCTCTACATT TACCGTGAGA CGATTGTGCT GTCCCTTGTG GGAATCGTAC TTGGTCTGAT	18060
AGCTGGTTTC TATTTACACC AATTTTGTAT TCAAATGATT TCGCCTGCGA CTATTCTCTT	18120
TTATCCGCAG GTAGGCTGGG AAGTCTATGT AATCCCAGTG GCAGCAGTAA GCATCATTTT	18180
GACCTTGCTT GGTTTCTTCG TCAATTATTA TCTGAGAAAG GTTGATATGT TAGAAGCCCT	18240
GAAATCTGTA GAGTAAGGTA GTTATTTTGA GCTGATTGAA CTCTATTTA CTAATATTCA	18300
AAAATCCTCC GTTTCAAAGA GCAGGGAACCT CTTTGTGACA GAGGATTTT TCTATAGGGC	18360
TTTAGCAGCT GCAATTGCGG CTTCGAAGTT TGGCTCAGAA TTGATATTAT CCACGTATTC	18420
AACGTAGCGA ATCGTATTGT CAGTATCGAG GACAAAGACT GCGCGTGCTA ATAGGTGCCA	18480
TTCGTTGATC AAGAGGGCAT AATCGCGCCC GAAAGAATGG TCAAAGTAGT CTGAAAGCAT	18540
AATGGCATTG TCAAGGCCTT CAGCACCGCA CCAACGTTTT TGAGCAAAAG GTAGGTCCAT	18600
TGAAACAGTC AATACGACCG TGTGTGCCAG TCCAGCCAAT TCTTCATTAA AACGACGTGT	18660
TTGAGTTGAG CAGATGCCTG TATCGATAGA AGGAACGACA CTCAAGACTT TTTCTTGCC	18720
ATCAAAATCA GCCAGAGATT TTTTAGAAAG ATCTGTTGTA GTAAGAGAAA AATCAAGCGC	18780
CTTGTCGCCG ACTTGTAGTT GTTTACCTGT AAAGCTCACA GGATTTCGA GAAAAGTTAC	18840
CATAGGATAC TCCAATCTTT TTTCTTCCAT TTTAGCTGAA ACAGTCGGA TTTTCCAATG	18900
ATTTGACCGG AAATATGGGC ATAGAAAAAA CGCCAGCTCA TGTGAGAATG ACGTTTTTCA	18960
TAGGTTTATT TTGCCAATCC TTCAGCAATC TTGTCAAGGT TGTATTTTAT CATGCTGTAG	19020
TAGCTGTGCG CTTCTTTACC TTGTTCTGCG ATAGAGTCAG TAAAGATTTG AGCGTAGATT	19080
GGGATGTTTG TGTCTTGAGA AACAGTTTTC ATTGGACGGT CATCCACACT TGATTCTACA	19140
AAGAGTGATG GAACTTTTGT TTGGCGAAGT TTTTCAACCA AGGTCCTGAT TTGTTTCAGGA	19200
GTTCCTTCTT CTTCAGTATT GATTTCCCAG ATGTAAGCAC TTGGGACACC ATAGGCTTTA	19260
GAGAAGTATT TGAATGCTCC TTCGCTGGTT ACAATGAGTT TCTTTTCAGC AGGGATCTTA	19320
TTAAATTTAT CCTTACTTTC TTTATCAAGT TTGTCTAACT TATCAGTATA TTCTTTGAGA	19380
TTTTTTTCAT AGAATCTTTT ATTGTTAGGG TCTTTGGCGC TCAATTGTTT GCGGATATTT	19440
TTAGCAAAAA TAATACCGTT TTCAAGGTTA AGCCAAGCGT GTGGGTCTTC TTTTCCTTTT	19500
TCATTTTGAC CTTCAAGGTA GATAACATCA ACGCCGTCGC TGAAGTCGAA GTAGTCTTTG	19560
TTTTTCAGTT TCTTGGCATT TTCTACCAAT TTTGTAAACC AAGCATTGCC ACCTGTTTCA	19620
AGGTTGATAC CGTTATAGAA AATCAAATTA GCCTCAGAAG TTTTCTTAAC GTCTTCAGGA	19680

370

AGTGGTTCGT ATTCGTGTGG GTCTTGCCCA ATCGGAACGA TACTATGAAG GTC AATTTTGG	19740
TCACCAGCAA TATTTT TAGT AATATCAGCG ATGATTGAGT TTGTAGCAAC AACTTTT TAGT	19800
TTT TGACCAG AAGTTGTATC TTTT TTTCCG CTAGCACATG CTACAAGAAT GATTGCAGAA	19860
AGAAAGAGAA CGAGTAATGT ACCTAATTTT TTCATTAGAT CCTCCAATTT ATTAGGGCTT	19920
TGCCCC TTAT TTTAACAAAT GTTTATTTT CAGTTTCAA TATCGTTGTT TGGGAGCGAT	19980
AAAGAAGCTA ATGAGAAAGA AACTAGCAGC TGTAAGCAGC ATACTAGAAC CTGCCGCAAC	20040
ATTAAACTA TAGCCAATAA AGAGTCCCAA AACTGAAGCA GTAGCTCCGA AGGTTGAGGA	20100
AAGGAAATC ATACTTTTCA GACTATTAGC ATACAGATAA GCAGTTGCAG CTGGGGTAAT	20160
CAGCATGGCT ACAATCAGGA TAGTTCCGAC ACTTTGCATG GCTGTCACAG ACACGAGAGT	20220
CAGGAGTACC ATGAGAAGGT AGTGATAGAA ATTGACAGGC ATTCCCATGG CTTTAGCCAA	20280
GAGTTCATCA AAGGAAGTTA TCAAGAGTTG CTTGAAGAAA ATCCAGATTA ACAAGAGGAT	20340
AGCTGCCCC ACACCCATAG TAATAACAT ATCCGTATCT TGGACGGCCA GGATATTACC	20400
AAAAAGGATA TGGAAAAGGT CAGTTGAACT TTTAGCGACA CCAATCAAGA TGATACCGAG	20460
GGCTAAGAAA GAAGAAAAGG TAATGCCGAT GGCGGTATCG CTTTGTATAA TCGAGTTTCC	20520
TTTGATGTAG GTAATGATGA TGGCAGCTAG CAATCCAAAG ACAATGGCTC CGATAAAGAA	20580
GTCAAGGCC AAGATGAAGG ATAGGGCTAC ACCTGGTAAG ACAGCATGTG AAATGGCATC	20640
TCCCATGAGT GACATCCCGC GTAGAATAAT GAAACATCCC ACAGCTCCAG CTACAATCCC	20700
GACGACAATA GCTGTTATCA AGGCATTTTG TAGGAAATGG AATTTTGTCA ATCCATCGAT	20760
AAATTCTGCA ATCATAGGTC ACCTCCATG AAAAAGAGTT GATTACCGTA AGCTTCTTTT	20820
AGATTGGTTT CCGTAAAAGT TTCTTTTGTT GGACCAAAG CAATCACTTC TCGATTGACA	20880
AGTAAGACTT GATCGAAGTA GTGGGGAATC TTGCTGAGGT CGTGGTGAAC GATGAGAACC	20940
GTCTTCCCAG CTTTTTCAA ATCTCTCAGC GTATTCATGA TGATTTCTC ACTGACAGAG	21000
TCAATCCCAG CAAAGGGTTC ATCCAAGAGG ATATAGTCGG CTTCTGTCAC CAAACATCTG	21060
GCAATCAAGA CCCGCTGGAA TTGACCTCCA GACAGTTGAC TAATTGACG TTCAGCGTAG	21120
TCAGCTAGGC CGACGATTTC AAGGGCCTCT TGCACTTTCT TCCAATGTTT AGCCTTTAAA	21180
CTTCGAAAGA GAGGAATAGA GGGAAATAGT CCTAACGAGA CGCATTCCTT GACCTTGATG	21240
GGAAAGTTGT AGTCGATATT GATTTTGT TCGACATAGG CAATTCGGTG TAAGGATTTT	21300
TTAACTTCCT TGTCATCGAG AAATGCCTGA CCTTGATGTG GGATAATTCC CAACATACCT	21360
TTTAATAGTG TTGATTTCCC AGCGCCGTTT GGACCAATGA TGCCGGTAAT TGTGGTCCA	21420
TGGAGCACTA GTGAAATATC CTTAAGTGCC AACGTTTCTT TGTAGGAGAC ACTGAGGTTT	21480

371

TCGATACGTA TCATAAACTT GTATTCCTCC TGTCTCTTAA TATACATTAA AAAAAAAATT	21540
AAGTCAAGTT AATTTTGTAA AAAATTAAAA TAATAACTGA AAAATAGATT CTAAAGATAA	21600
CTTTCAGGAT AAATTTCTAA ATTATAAAAC GCATAGTATC AAGTGTAATA AACTTGGAAT	21660
TATGCGTTTT ATCATGGAAA GATTTTTTAT AATAGCTAAA AAATAA	21706

(2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5171 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GATCCCCAGG AAAAACCAGG GTTTTCCCAA TCAATCGTTA CTGTCATATT CCACTCCTTA	60
TTCTAAAAAC CTATTTCTTA TATTCTACAC TATTTTCTA AAATAGCAAG TATATTTTGT	120
AATTTTCAGA AAATTTCTCC AATAAAAACC AACTCTTAGA ACTGATTCTT CATTTCACTT	180
ATTTATCTTC AGTAACTACT TCCTGAAGAT AAGCGTCAAA AACTTCTTCA TCTGAAATCG	240
TGTCAGAAAT GAAGCTTCCA TTGCTAGTGC GTTCTGACAA GTTCAAGTCT TGCAATCGGC	300
TTTCATAGAT TGTTCCCTTA TTGGATTGGA CAAGCAGAGT TTGGTCGTTT ACATCCACTT	360
CCGTACTGAA GAAATCGCCA ACAAATCCTT GCTCTGCAAC TGCTCCTGCC AAGAAGACAC	420
GATGCGGTTT GTTTTTC AAC TCACGCAAGA CTGTGAATCC TCGTTTGGCA CGGCTGGTTG	480
CTAGAATTTT CTCAATGGAA ACACGTTTCA AGCTTCCACG CTGGGTCAAG AGGTAGAAGG	540
ACGAAGTATT ACAGATAAAG CCAGATTGGA GGACATCATC TTCTTTCAA TTCATAGCCT	600
TGACACCTGC TGCCTTAGCA CCGACAACCG GAACCTCTTC GATATTGAAA CGCAGGGCAT	660
AACCATTTTG ACTAACCAAG ACAACATCAT CTAGTTTAAT CGGAGCCACT GCTACAATCT	720
GATCTGTATC GTCTTTGAGC TTAGCATACT TGACAGACTT AGATCTATAG GTCCGCCATG	780
GAGTGAATTC TTTTCGCTCT ACCCGTTTGA TTTGACCAAG GCGAGTCACT GCAAAGTAGG	840
TTGTGCGATC GTCAAACCTGA TCCAGTACTT CCACATAAAG GATTCTTCA TTCGTTTCAA	900
AGTTTGTGAT GGTTTGGCTC AGATGCTCTC CGATGTCCTT CCAACGAATA TCTGCCAACT	960
CATGGATTGG TCTGTAGATG ACATTTCCAA GACTTGTGAA CATCAAGAGG TGCTGGGTTG	1020
TCTTGGCAGA TTGAACAAAA ATCAAACGGT CATCATCACG CTTGCCAATT TCTTCCAAGG	1080
TGGAAGCCGC AAAGGAACGT GGAAGGTAC GCTTGATGTA ACCTGCCTTG GTCACGCTGA	1140

372

CGTAGGTATC TTCCTCAGCG ATAAGACTAG CTGTATCAAT CTCAATTGCT TTCGCAGTGT	1200
CTTCTAAAGA ACTCAAACGA GGAGTTGCAA ATTTCTTCTT GACCTCACGA AGTTCTTTCT	1260
TCATGAGATT GTACATAGTC CTTTCATCAC CGATAATAGC CGCCAGCATA GCAATCTTCT	1320
CACGAAGCTC TGCTTCTTCT TCCTGCAAGA CAACCACATC GGTATTGGTC AAACGGTACA	1380
GTTGCAAAAGT TACGATAGCC TCAGCCTGTT CTTCCTGAAA ATCATAGCTA ACTTTGAGGT	1440
TTTCCTTGGC GTCCGCCCTTA TTCTCAGAAG CACGGATAAG AGCAATGACT TCATCCAAAA	1500
TCGAAATCAC ACGAATCAAA CCTTCGACGA TATGGAGACG TTTCTCAGCC TTTTCTTTGT	1560
CAAAGCGTGA ACGCGCCAAA ATCACTTCTC GACGGTGAGC GATATAGCTA GACAGGATTG	1620
GAACAATCCC AACCTGACGA GGTGTGAAAT TGTCAATCGC CACCATATTA AAGTTGTAGT	1680
TGATTTGTAG GTCGGTGATC TTAAATAAGT AGTTGAGAAC AAGCTCAGTA TTAGCGTCTT	1740
TCTTAAGTTC GATAGCGATA CGAAGACCAT CACGGTCAGA CTCATCACGA ACCTCAGCAA	1800
TCCCAGCTAC CTTGTATTATTA ACACGAACAT CATCGATTTT CTTGACTAGA TTGGCCTTAT	1860
TGATTTTATA AGGAATCTCA ATAATAACGA TTTGTTCCTT ACCACCTTTT AGCTTTTCAA	1920
TTTCAGTCTT GGAACGAACA ACCACGCGCC CTTTCCCAGT CTCATAAGCT TTCTTGATT	1980
CATCAGGACC CTGAATAATA GCCCCTGTAG GGAAGTCTGG TCCAGGCAAG AATTCATGA	2040
GTTTATCAAT CTTTGCAAGT GGGTGGTCAA TCATGTAAAC TGCAGCATCT ATGACCTCAG	2100
CTAAATTATG GGGAGGAATG TCTGTGGCAT AACCAGCCGA AATCCCAGTC GAACCATGA	2160
CCAAGAGGTT TGGAAAGGCT GCTGGCAAGA CCGTTGGTTC TTTCTCCGTA TCGTCAAAGT	2220
TCCATGCAAA AGGAACTGTC TTTTCTCGA TATCCTGAAG AAGGTAGCCT GCAATTTAG	2280
ACAAACGTGC CTCAGTATAA CGCATAGCCG CAGGAGGATC TCCGTCCATA GAACCGTTAT	2340
TACCGTGCAT TTCAACTAGA ATCTCAGCAT TTTTCCAGTT CTGTGACATA CGAACCATGG	2400
CATCATAGAT AGAAGAATCC CCGTGTGGGT GGAAATTCCC CATGATGTTC CCGACTGACT	2460
TGGCCGACTT ACGGTAGCTC TTGTCAAAAG TATTGCTATC CTTATTGATA GAATAAAGAA	2520
TACGGCGCTG AACCAGGCTC AACCACATC GAATATCTGG CAAAGCCCGG TCTTGAATAA	2580
TGTACTTGGA GTAGCGACCA AAGCGCTCTC CCATGATGTC CTCCAGGGAC ATGTTTGGAA	2640
TGTTAGACAT AAGATACAAA GCCCATAAAA TACCAAGTGA AAATAGAAAA TTCTTGAAGT	2700
AAGCAAACCTC ACAAGAGAAT TTATCTTTTT CACACAGTAT CTAGGGCGTG TTCAACTCCT	2760
TTCAAAGAAT GTAGAGTAGG TTTTATGCA GTAAAAGATA TTTTACGGGA ATTCTCCCG	2820
TGTTCAAGTA CGATAAGTAA CCAAACATC CTGTTTGTAT TTTTCAATAT GAAAATCTGG	2880
TTTTCCAAAA TTAGTCTTAG TTTGTGCTT AGCCGCTCCC TTAAGCGCCT CTTTGAGATA	2940

373

AGCACTCATA GCAGATTCTT CATTAATAAT CCTGCAATTT TTTCAAACCA AGATTTTCAA	3000
ACTGCTTTTT CACATAGTCA TTCACATCCG ACTCTAATTT CCAGTTTACT AACATATTAT	3060
TTTCTTTTCAT TAAAACACTG TCGTTTCTTC TAGCGTAAAC TTGACATTAT CTTCAATCCA	3120
TTTACGGCGT GGTTCACCT TATCTCCCAT GAGAACATTG ACGCGGCGTT CGGCGCGCGC	3180
TAAATCTTCA ATTGTGACAC GGATGAGGT ACGTGTTCCT GGGTTCATGG TTGTTTCCCA	3240
GAGCTGGTCC GCATTCATCT CACCAAGTCC TTGTATCGT TGGAGGGTAG CGCCTTTACC	3300
GAATGTTTA CGGAGTTCTT CTAGTTCTCC GTCCGTCCA GCGTAGGCCA CTTCTTCTTT	3360
CTTGCCTTTA CCTTGGACA TCTGTAAAG AGGTGGGAGG GCAATATAGA CATGACCTGC	3420
CTCGACTAGC GGACGCATGT AACGGTAGAA AAATGTCAAG AGCAAGGTCT GGATATGGGC	3480
ACCGTCGGTA TCCGCATCGG TCATGATAAT GATCTTATCA TAGTTGGCAT CTTCAATAGA	3540
GAAGTCTGCT CCAACACCCG CACCAATGGT ATAAATCATG GTATTGATCT CTTCATTTTT	3600
GAGGATATCC GCCATCTTGG CCTTGGCTGT ATTGACAACC TTACCACGAA GAGGTAGAAT	3660
AGCCTGGAAC TTGCGGTCAC GACCTTGTTC GGCAGAACCA CCGGCAGAGT CCCCCTCAAC	3720
TAGATAGAGT TCATTCCTAG CAGGATCTT AGATTGGGCT GGGGTCAATT TCCCAGACAA	3780
CAAGCCCTTA TCTTCTTGT TTTCTTCCC ATTCGGCTC TCATCACGCG CCTTACGTGC	3840
TGCTTCACGA GCATCACGGG CCTTGATAGC CTGCGGATG AGGTAGAAG CTAATTCCCC	3900
ATTTTCCATA AGGAAAAGG TCAACTTATC AGCCACTATT CCATCCACAA CTGGGCGAGC	3960
TAGGGGGCTT CCTAGTTTAT CCTTGGTCTG TCCTTCAAAC TGCAAGTGT CTTCAGGAAC	4020
TAAGATAGAA AGAACGGCCG CTAGTCCCTC ACGATAGTCT GAACCTTCAA GGTTTTTATC	4080
TTTTTCCTTG AGAAGACCTG TTTTACGTGC ATAGTCATTC ATGACCTTGG TAATGGCAGA	4140
CTTGAGTCCT GTCTCGTGG TTCCACCGTC CTTGGTGCGA ACGTTATTGA CAAAAGATAG	4200
AATGTTATCT GAGAATCCGT CATTGTACTG GAGGGCTACT TCCACTTGAA AACCATTGTC	4260
TTCCCTTCA AAGTAAAGAA CTGGCGTCAA GATTTCCCTTA TCTTCGTTGA GATAAGAAAC	4320
AAAATCTGT ACTCCATTCT CATAGTGGA CTCAATCGCT TCATTTGTTC GCTTGTCCGT	4380
TAAAGACAAG GTCACATTTT TCAAGAGAAA GGCTGATTCA TTAAGGCGCT CTGAAATGGT	4440
ATTGTACTTG AAATCTGTG TAGAAAATAT AGTCGCGTCA GGCATAAAAG TAACTTTGGT	4500
GCCTGTTTTA GACTTGGGTG CTGTACCGAT TTTCTTCAA GTCTGACAG GTTTTCCACC	4560
ATTTTCGAAA CGTTGCTTGT AAATGCGCC ATCACGGGTA ATTTCAACTT CTAACCAGCT	4620
AGAAAGGGCG TTAACAACGG AAGAACCAC TCCGTGAAGT CCACCTGATG TCTTATAGCC	4680

374

ACCTTGACCG AATTTCCTC CGGCATGAAG AATGGTAAAG ATAACCTCAA CAGTTGGAAT	4740
TCCCATAGCG TGCATACcTG TCGGCATCCC ACGTCCATGG TCTTGAACCG TTAGACTACC	4800
GTCTTTATTG ATAGTTACAT CAATACGATC ACCAAACCCA GACAAGGCTT CATCGACTGC	4860
ATTATCAACG ATTTCCCAAA CTAGGTGATG AAGACCAGCG CCATCGGTCG ATCCAATATA	4920
CATCCCTGGA CGTTTTCGGA CCGCATCCAA CCCTTCTAGC ACCTGAATAG CATCATCATT	4980
ATAATTGTTA ATATTGATTT CCTTTTTTGA CACAAGGAAC CTCCTATTCG TFCATCTTTA	5040
CTATCTACA GGTTTTCCAA GGATTTTGCA AAATTTTCTT TTCTCCGATG TGACAATTTT	5100
AGCAGAGATT CTCTGCTTTT CTTTCCCAAT TCATGATATA ATAGGAGTAT GATTACAATA	5160
GTTTTATTAA TCCTAGCCTA TCTGCTGGGT TCGATTCCAT CTGGTCTCTG GATTGGACAA	5220
GTATCTTTTC AAATCAATCT ACGCGAGCAT GGTTCCTGTA ACGTGGAAC GACCAACACC	5280
TTCCGCATTT TAGGTAAGAA AGCTGGTATG GCAACCTTTG TGATTGACTT TTTCAAAGGA	5340
ACCCTAGCAA CGCTGCTTCC GATTATTTTT CATCTACAAG GCGTTTCTCC TCTCATCTTT	5400
GGACTTTTGG CTGTTATCGG CCATACCTTC CCTATCTTTG CAGGATTTAA AGGTGGTAAG	5460
GCTGTGCGAA CCAGTGTGG AGTGATTTTC GGATTTGCGC CTATCTTCTG TCTCTACCTT	5520
GCGATTATCT TCTTTGGAGC TCTCTATCTT GGCAGTATGA TTCACTGTC TAGTGTCACA	5580
GCACTGATTG CGGCTGTTAT CGGGGTCTG CTCTTTCCAC TTTTGGTTT TATCCTGAGT	5640
AACTATGACT CTCTCTTCAT CGCTATTATC TTAGCACTTG CTAGTTTGAT TATCATTCGT	5700
CATAAGGACA ATATAGCTCG TATCAAAAAT AAAACTGAAA ATTTGGTCCC TTGGGGATTG	5760
AACCTAACCC ATCAAGATCC TAAAAAATAA AATGCCAGTT CTGTACTGCC CCCAACAGT	5820
TAGACAAATA ATTTATCCAA AGGATTTAGT TCTGTACTGC ACAGGACTAA GTCCTTTTAG	5880
TTTACCTTA ATTCGTTTGT TGTGTAGTA ATCAATATAG TCTATAATGG CTTGTTCCAA	5940
TTGATTAAAGT GATTTAAATG TTTTCTCATA GCCATAAAC ATTTCCGATT TTAAATGCC	6000
AAAGAAAGAT TCCATCCTAC CGTTGTCTTG GCTGTGCCC TTACGTGACA TGGATGCTTG	6060
AATTCCTTA CTCTCTAGGA ACCGATGATA AGAATCGTGT TGGTATTGCC AGCCTTGGTC	6120
ACTATGGAGA ATCGTATTCT CGTAGTGCTT CTCTGTGAAT GCCTGTCCA A	6171

(2) INFORMATION FOR SEQ ID NO: 38:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18475 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

TATTACAAAT AAAAAACGG AGGAGTGCTT TATGAAAGCC TATACTTATG TTAAACCAGC	60
ACTTGCTTCT TTTGTGATG TAGACAAACC AGTTATTCGC AAGCCAACAG ACGCTATTGT	120
GCGTATTGTA AAAACCACTA TTTGTGGAAC AGACCTCCAT ATTATCAAAG GGGATGTTCC	180
TACTTGCCAA AGTGGTACCA TTCTTGGCCA CGAAGGGATT GGGATTGTTG AAGAAGTTGG	240
GGAAGGAGTT TCCAACCTCA AAAAAGGTGA CAAGGTCTTG ATTTCTTGCG TCTGTGCCTG	300
TGGTAAATGC TACTACTGTA AAAAAGGAAT TTATGCTCAC TGTGAAGACG AAGGGGGCTG	360
GATTTTCGGT CACTTGATTG ATGGTATGCA GGCTGAATAT CTACGTGTCC CTCATGCAGA	420
TAATACTCTT TACCATACTC CAGAAGACTT GTCAGATGAA GCTTTGGTTA TGCTGTCAGA	480
CATCTGCCT ACTGGATATG AAATTGGTGT CTTAAAAGGG AAAGTAGAAC CTGGTTGCAG	540
CGTAGCCATT ATTGGTTCAG GTCCAGTTGG ATTGGCTGCT CTTTAAACAG CCCAATTCTA	600
TTCAACAGCT AAATTGATTA TGGTAGACCT AGACGATAAC CGCTTGAAA CTGCCCTATC	660
ATTCGGTTCG ACTCATAAGG TTAATCTTC AGACCTGAA AAAGCCATTA AAGAAATTTA	720
TGATTTGACA GATGGTCGTG GTGTGGATGT CGCTATCGAA GCTGTTGGTA TTCCTGCAAC	780
ATTTGATTTT TGTCAAAGA TTATCGGTGT AGACGGAACG GTTGCCAACT GTGGTGTGCA	840
TGGTAAACCA GTTGAATTCG ATTTAGATAA ACTTTGGATT CGCAACATCA ATGTAACAAC	900
TGGTTTGTA TCTACAAATA CGACTCCACA ATTGTTGAAA GCACCTGAAA GTCATAAGAT	960
TGAACCGAA AAATTGGTAA CTCACTATTT CAACTCAGT GAAATTGAAA AAGCCTACGA	1020
AGTCTTCAGT AAGCAGCAG ACCACCATGC CATTAAGGTC ATTATCGAAA ACGATATCTC	1080
AGAAGCCTAA GTAGTAAAA TATTTTGTG CATAAGTAAA TAGAAATCA GTCATCCATC	1140
AGATGGCTGG ATTTTTATC AAAAAATTAA GAAATGAGCA TATTTCTTTC CTTGTCTGGC	1200
GGAATTGGTT ATAATATACG GTACAAAGGA ATGAATGAAT ATGTATCGTG TTATAGAAAT	1260
GTACGGAGAT TTTGAACCGT GGTGGTCTT AGAAGTTGG GAAGAAGATA TTGTAGCAAG	1320
TAGAAAATTT GACCAGTATT ATGATGCTCT CAAATACTAC AAAACTTGCT GGTTTAGATT	1380
GGAACAAGAA TCGCCTCTTT ATAAAAGTAG AAGCGACTTG ATGACCATT TTTGGGACCC	1440
GGAAGACCAA CGCTGGTGTG ATGAATGTGA TGAGTATTTA CAACAATACC ATTCTTTGGC	1500
TCTTTTGCAG GATGAGCAGG TTATCCCAGA CGAAAACTA CGCTCAGGCT ATGAAAAACA	1560
AACCAATCAG GAAAGGAATC GTTCTTGCCG TATGAAATTA AAATAGACAA AAGTAACTTT	1620
TTTGGAGTTG CTTTTTTAT TTTCTAACT CTTTGCGAAT AGTATAGGTG AGGAGGTAAG	1680

376

TATGGTTCAA GAAATTGCAC AAGAAATCAT TCGTTCAGCT CGGAAAAAAG GGACGCAGGA	1740
TATCTATTTT GTCCCTAAGT TAGACGCCTA TGAGCTTCAT ATGAGGGTAG GAGACGAGCG	1800
CTGTAAAATT GGTAGCTATG ATTTTGAAAA GTTTCAGCC GTTATCAGTC ACTTTAAGTT	1860
TGTGGCGGGT ATGAATGTGG GAGAAAAAAG ACGTAGTCAA CTGGGTTCCCT GTGATTATGC	1920
CTATGACCAT AAGATAGCGT CTCTACGTTT ATCTACTGTA GGCGATTATC GGGGGCATGA	1980
GAGTTTGGTT ATCCGTTTGT TGCACGATGA GGAGCAGGAC CTGCATTTTTT GGTTTCAGGA	2040
TATTGAAGAA TTAGGCAAGC AGTACAGGCA ACGGGGACTC TATCTTTTTG CTGGTCCGGT	2100
TGGGAGTGGT AAGACGACCT TGATGCATGA ATTGTCCAAG TCACTCTTTA AAGGACAGCA	2160
AGTTATGTCC ATCGAAGATC CTGTGAAAT CAAGCAGGAC GACATGCTTC AGTTGCAGTT	2220
GAACGAAGCA ATCGGCCTAA CCTATGAAAA TCTAATCAAA CTTTCCTTGC GTCATCGACC	2280
AGATCTCTTG ATTATCGGAG AAATTCGTGA CAGCGAGACG GCGCGTGCAG TGGTCAGAGC	2340
TAGTTTGACA GGTGCGACAG TCTTTTCAAC CATTACGCC AAGAGTATCC GAGGTGTTTA	2400
TGAGCGTCTG CTGGAGTTGG GTGTGAGTGA AGAAGAATTG GCAGTTGTTC TGCAAGGAGT	2460
CTGCTACCAG AGATTAATCG GGGGAGGAGG AATCGTTGAC TTTGCAAGCA GAGATTATCA	2520
AGAACACCAA GCAGCCAAGT GGAATGAGCA AATTGACCAG CTTCTTAAAG ATGGACATAT	2580
CACAAGTCTT CAGGCTGAGA CGGAAAAAAT TAGCTACAGC TAAGCAAAAA AATATCATCA	2640
CCCTATTTAA CAATCTCTTT TCTAGCGGTT TTCATCTGGT GGAGACTATC TCCTTTTATG	2700
ATAGGAGTGC TTTGTGGAC AAGCAGTGTG TGACCCAGAT GCGTGTGGGC TTGTCTCAGG	2760
GGAAATCATT CTCAGAAATG ATGGAAAGTT TGGGATGTTT AAGTGCTATT GTCACTCAGT	2820
TATCCCTAGC TGAAGTTCAT GGCAATCTCC ACCTGAGTTT GGGAAAGATA GAAGAATATC	2880
TGGACAATCT GGCTAAGGTC AAGAAAAAAT TGATTGAAGT AGCGACCTAT CCCTTGATTT	2940
TGCTGGGTTT TCTTCTCTTA ATTATGCTGG GGCTACGGAA TTACCTGCTC CCACAACTGG	3000
ATAGTAGCAA TATTGCCACC CAAATTATCG GTAATCTGCC CCAAATTTTT CTAGGCATGG	3060
TAGGGCTTGT TTCCGTGCTT GCCCTTTTAG CACTCACTTT TTATAAAAGA AGTTCTAAGA	3120
TGAGTGTCTT TTCTATCTTA GCACGCCTTC CCTTTATTGG AATCTTTGTG CAGACCTACT	3180
TGACAGCCTA TTAGTCACGT GAATGGGGGA ATATGATTTT ACAGGGAATG GAGTTGACGC	3240
AGATTTTCA AATGATGCAG GAACAAGGTT CCCAGCTCTT TAAAGAAGTC GGTCAAGATC	3300
TGGCTCAAAC CCTGAAAAAT GGCCGTGAAT TTTCTCAGAC GATAGGAACC TATCCTTTCT	3360
TTAGGAAGGA ATTGAGTCTC ATCATAGAGT ATGGGGAAGT TAAGTCCAAG CTGGGTAGTG	3420
AGTTGGAAT CTATGCTGAA AAAACTTGGG AAGCCTTTTT TACCCGAGTC AACCACCA	3480

TGAATTTGGT GCAGCCACTG GTTTTATCT TTGTGGCACT GATTATCGTT TTA	3540
CGGCAATGCT CATGCCCATG TATCAAAATA TGGAGGTAAA TTTTAAAT GAAAAAATG	3600
ATGACATTCT TGAAAAAGC TAAGGTAAA GCTTTTACAT TGGTGGAGAT GTTGGTGGTC	3660
TTGCTGATTA TCAGCGTGCT TTTCTTGCTC TTTGTACCTA ATCTGACCAA GCAAAAAGAA	3720
GCAGTCAATG ACAAAGGAAA AGCAGCTGTT GTTAAGGTGG TGGAAAGCCA GGCAGAACTT	3780
TATAGCTTAG AAAAGAATGA AGATGCTAGC CTAAGAAAGT TACAAGCAGA TGGACGCATC	3840
ACGGAAGAAC AGGCTAAAGC TTATAAGAA TACAATGATA AAAATGGAGG AGCAAAATCGT	3900
AAAGTCAATG ATTAAGGCCT TTACCATGCT GGAAAGTCTC TTGGTTTGG GACTTGTGAG	3960
TATCCTTGCC TTGGGCTTGT CCGGCTCTGT CCAGTCCACT TTTTCAGCGG TAGAGGAACA	4020
GATTTTCTTT ATGGAGTTTG AAGAACTCTA TCGGGAAACC CAAAAACGCA GTGTAGCCAG	4080
TCAGCAAAAG ACTAGTCTGA ACTTAGATGG GCAGACGCTT AGCAATGGCA GTCAAAAGTT	4140
GCCAGTCCCT AAAGGAATTC AGGCCCATC AGGCCAAAGT ATTACATTTG ACCGAGCTGG	4200
GGGCAATTCG TCCCTGGCTA AGGTGAATT TCAGACCAAGT AAAGGAGCGA TTCGCTATCA	4260
ATTATATCTA GGAATGGAA AAATTAACG CATTAAGGAA ACAAAAAATT AGGGCAGTGA	4320
TTTTACTGGA AGCAGTAGTC GCTCTAGCTA TCTTTGCCAG CATTGCGACC CTCCTTTTGG	4380
GACAAATCA AAAAAATAGG CAAGAGGAAG CAAAAATCTT GCAAAAGGAA GAAGTCTTGA	4440
GGGTAGCTAA GATGGCCCTG CAGACGGGGC AAAATCAGGT AAGCATCAAC GGAGTTGAGA	4500
TTCAGGTATT TTCTAGTGAA AAAGGATTGG AGGTCTACCA TGGTTCAGAA CAGTTGTGG	4560
CAATCAAGA GCCATAAGGT CAAGGCTTTT ACCTTGTTAG AATCCCTGCT TGCCCTCATT	4620
GTCATCAGTG GGGGATTACT CCTTTTCAA GCTATGAGTC AGCTCCTCAT TTCAGAAGTT	4680
CGTACCAGC AACAAAGCGA GCAAAAGGAG TGGCTCTTGT TTGTGGACCA ACTTGAGGTA	4740
GAATTAGACC GTTCGCAGTT CGAAAAAGTA GAAGGCAATC GCCTATACAT GAAGCAAGAT	4800
GGCAAGGACA TCGCCATCGG TAAGTCAAAG TCAGATGATT TCCGTAAC GAATGCTCGT	4860
GGTCGAGGTT ATCAGCCTAT GGTATATGGA CTCAAATCTG TACGGATTAC AGAGGACAAT	4920
CAACTGGTTC GCTTTCATTT CCAGTTCCAA AAAGGCTTAG AAAGGGAGTT CATCTATCGT	4980
GTGGAAGGAG AAAAAAGTTA AGGCAGGTGT TCTCCTCTAC GCAGTCACCA TAGCAGCCAT	5040
CTTAGTCTT TTGTGCAAT TTTATTTGAA CCGACAAGTC GCCACTATC AAGACTATGC	5100
TTTGAATAAA GAAAAATTGG TTGCTTTTGC TATGGCTAAA CGAACCAAG ATAAGGTTGA	5160
GCAAGAAAGT GGGGAACAGT TTTTAAATCT AGGTCAGGTA AGCTATCAAA ACAAGAAAC	5220

378

TGGCTTAGTG	ACGAGGGTTC	GTACGGATAA	GAGCCAATAT	GAGTTTCTGT	TTCCTTCAGT	5280
CAAAATCAAA	GAAGAGAAAA	GAGATAAAAA	GGAAGAGGTA	GCGACCGATT	CAAGCGAAAA	5340
AGTGGAGAAG	AAAAAATCAG	AAGAGAAGCC	TGAAAAGAAA	GAGAATTCAT	AGTCAATTCA	5400
ACTATAATGC	GTTGAATCCA	GAATAGTCCA	CTGTAGTTTC	TAGAAAATTG	CTGGAAATGG	5460
ATGTTAAGCT	CCAATTCATT	TGTTTATATC	TTATTTTCAGT	TTACTATACT	TTGTGCTAAA	5520
TTAAAGATAT	GAAACATGAT	TTTAACCACA	AAGCAGAAAC	TTTCGATTCC	CCTAAAAATA	5580
TCTTCCTCGC	AAACTTGGTA	TGTCAAGCAG	CCGAGAAACA	GATTGATCTT	CTATCAGACA	5640
AAGAAATTTT	AGATTTTCGGT	GGTGGCACGG	GTCTATTAGC	CTTGCCCTTA	ACCCCTAGCC	5700
AAGCAGGCTA	AGTCAGTCAC	TCTTGTAGAC	ATTTCTGAGA	AAATGTTGGA	GCAAGCTCGT	5760
TTGAAAGTGG	AGCAGCAAGC	AATCAAGAAT	ATCCAGTTTT	TGGAGCAAGA	TTTACCAGAA	5820
AATCCCTTGG	AGAAAGAGTT	TGATTGCCTT	GCTGTTAGTC	GGGTTCTTCA	TCATATGCCT	5880
GATTTGGATG	CGGCTCTCTC	ACTGTTTCAT	CAACATTTGA	AGGAAGATGG	GAAACTCATC	5940
ATTGCTGATT	TTACCAAGAC	AGAAGCTAAT	CATCATGGAT	TTGATTTAGC	TGAACTGGAA	6000
AACAAGCTAA	TTGAGCATGG	TTTTTCATCT	GTGCATAGTC	AGATTCTCTA	TAGTGCTGAA	6060
GACCTGTTTC	AAGGAAATCA	CTCAGAATTC	TTTTTAATAG	TAGCCCCAAA	ATCACTCGCC	6120
TAGTCAGGGA	GTGATTTTTTC	TATAAGGATG	GAAAAAAGAA	GGGAAATTTG	GTAAGATAGG	6180
AATATGGATT	TTGAAAAAAT	TGAACAAGCT	TATACCTATT	TACTAGAGAA	TGTCCAAGTC	6240
ATCCAAAGTG	ATTTGGCGAC	CAACTTTTAT	GACGCCTTGG	TGGAGCAAAA	TAGCATCTAT	6300
CTGGATGGTG	AAACTGAGCT	AAACCAGGTC	AAGGAGAACA	ATCAAACCCT	TAAGCGTTTA	6360
GCACTACGCA	AAGAAGAATG	GCTCAAGACC	TACCAAGTTT	TCTTGATGAA	GGCTGGGCAA	6420
ACAGAACCCT	TGCAGGCCAA	TCACCAGTTT	ACACCGGATG	CTATTGCTTT	GCTTTTGGTG	6480
TTTATTGTGG	AAGAGTTGTT	TAAAGAGGAG	GAAATTACTA	TCCTCGAAAT	GGGTTCTGGG	6540
ATGGGAATTC	TAGGCGCTAT	TTTCTTGACC	TCGCTTACTA	AAAAGGTGGA	TTACTTGGA	6600
ATGGAAGTGG	ATGATTTGCT	GATTGATCTG	GCAGCTAGCA	TGGCAGATGT	AATTGGTTTG	6660
CAGGCTGGCT	TTGTCCAAGG	AGATGCCGTT	CGCCCAAAA	TGCTCAAAGA	AAGCGATGTG	6720
GTCATCAGTG	ACTTGCCTGT	CGGCTATTAT	CCTGATGATG	CCGTTGCGTC	GCGCCATCAA	6780
GTTGCTTCTA	GCCAAGAACA	TACTTACGCC	CATCACTTGC	TCATGGAACA	AGGGCTTAAG	6840
TACCTCAAGT	CAGACGATA	CGCTATTTTT	CTAGCTCCGA	GTGATTTGTT	GACCACTCCT	6900
CAAAGTGATT	TGTTAAAAGA	ATGGCTGAAA	GAAGAGGCGA	GTCTGGTTGC	TATGATTAGT	6960
CTGCCTGAAA	ATCTCTTTGC	TAATGCCAAA	CAATCTAAGA	CTATTTTAT	CTTACAGAAG	7020

379

AAAAATGAAA TAGCAGTAGA GCCTTTTGTT TATCCACTTG CTAGCTTGCA AGATGCAAGT	7080
GTTTTAATGA AATTAAAGA AAATTTTCAA AAATGGACTC AAGGTACTGA AATATAAAAT	7140
AGATTTTGTT ATAATAGTTG AAAACGCTTA AAAAGGGGTA TCATGTTATG ACAAAAACAA	7200
TTGCAATCAA TGCAGGAAGT TCAAGTTTGA AATGGCAATT ATACTTAATG CCAGAAGAAA	7260
AAGTATTGGC GAAAGGTTTG ATTGAACGTA TCGGTTTGAA AGATTCAATT TCAACTGTAA	7320
AATTTGACGG CCGTTCTGAA CAACAAATTT TGGATATTGA AAATCATATA CAAGCCGTTA	7380
AAATTTTATT GGATGACTTG ATTCGTTTCG ATATTATCAA GGCTTATGAC GAGATTACAG	7440
GTGTTGGACA TCGTGTGTT GCTGGTGGAG AATATTTCAA AGAATCAACA GTTGTGAGG	7500
GAGATGTTTT AGAAAAAGTT GAAGAGTTGA GTTTGTTGGC TCCTCTACAC AACCCGGCCA	7560
ATGCAGCAGG TGTTCTGTGCC TTCAAGGAAT TGTGTCAGCA CATTACCAGT GTAGTTGTTT	7620
TTGATACTTC CTTCCACACA AGTATGCCAG AGAAAGCTTA TCGCTACCTT CTACCAACAA	7680
AATATTACAC AGAAAACAAG GTTCGTAAAT ACGGTGCTCA TGGTACAAGT CACCAGTTTG	7740
TAGCAGGAGA AGCTGCAAAA CTCTTGGGAC GTCCATTAGA AGACTTGAAG TTAATTACCT	7800
GTATATTGG TAACGGAGGC TCAATTACAG CTGTGAAAGC CGGCAATCT GTAGACACTT	7860
CTATGGGGTT CACTCCTCTT GGTGGTATTA TGATGGGAAC GCGTACAGGG GATATTGATC	7920
CAGCTATCAT TCCTTATTTA ATGCAATATA CAGAGGATTT TAACACACCA GAAGATATCA	7980
GTCTGTCTCT TAACCGTGAA TCAGGTCTTT TGGGAGTTTC TGCTAATCT AGCGATATGC	8040
GCGATATAGA AGCAGCTGTA GCAGAAGGGA ATCACGAGGC TAGCTTGGCT TATGAAATGT	8100
ATGTTGACCG TATCCAAAAA CATATCGGTC AGTACCTTGC AGTGCTAAAT GGAGCAGATG	8160
CCATTGTTTT CACAGCAGGT GTCGGTGAAA ATGCAGAGAG TTTCCGTCGT GATGTAATCT	8220
CAGGGATTTT GTGGTTTGGT TGTGATGTTG ATGATGAAAA GAATGTCTTT GCGGTTACAG	8280
GAGACATCTC AACAGAGGCA GCTAAAATCC GTGTCTTGGT TATTCCAACA GATGAAGAAT	8340
TAGTCATTGC CCGTGACGTT GAACGCTTGA AAAAATAAGT GAAACTAAAA AAATATTCAA	8400
TACAAGGAGT TGGGAAAGTT ATTTTTCAG CTCTTTTTC TGATGAAATT GTCCAAAACC	8460
TTGCTATGAT TGGCTTTTTT GAAAAATATG GTATAATAGT AGTAATTTAA TAGATGGAGT	8520
TGAGTTTGA AGAAAACTT TCGTGTAATA AGAGAGAAAG ATTTTAAGGC GATTTTCAAG	8580
GAGGGGACAA GTTTTGCTAA TCGCAATTT GTGGTCTACC AATTAGAAAA CCAGAAAAAC	8640
CGTTTTCGAG TAGGTCTATC AGTTAGCAAA AAAGTGGGGA ATGCCGTCAC TAGAAATCAA	8700
ATTAAGCGAC GGATTCCGCA TATTATCCAG AATGCAAAAG GGAGTCTGGT AGAAGATGTC	8760

380

GACTTTGTTG TCATTGCTCG AAAAGGAGTC GAAACCTTGG GATACGCAGA GATGGAGAAA	8820
AATCTACTCC ATGTATTAAA ATTATCAAAG ATTTACCGGG AAGGAAATGG GAGTGAAAAA	8880
GAAACTAAAG TTGACTAGTT TGCTAGGACT GTCTCTGTTA ATCATGACAG CCTGTGCGAC	8940
TAATGGGGTA ACTAGCGATA TTACAGCCGA ATCGGCTGAT TTTTGGAGTA AATTGGTTTA	9000
CTTCTTTGCG GAAATCATTC GCTTTTATC GTTTGATATT AGTATCGGAG TGGGGATTAT	9060
TCTCTTTACG GTCTTGATTG GTACAGTCCT CTTGCCAGTC TTTCAGGTGC AAATGGTGGC	9120
TTCTAGGAAA ATGCAGGAAG CTCAGCCACG CATTAAGGCG CTTCGAGAAC AATATCCAGG	9180
TCGAGATATG GAAAGCAGAA CCAAACTAGA GCAGGAAATG CGTAAAGTAT TAAAGAAAT	9240
GGGTGTCAGA CAGTCAGACT CTCCTTGCCG GATTTTGATT CAGATGCCGG TTATTTTGGC	9300
CCTGTTCCAA GCCCTATCAA GAGTTGACTT TTAAAGACA GGTCATTTCT TATGGATTAA	9360
CCTTGGTAGT GTGGATACAA CCCTTGTCT TCCGATTTTA GCAGCAGTAT TCACCTTTT	9420
AAGTACTTGG TTGTCCAACA AAGCTTTGTC TGAGCGAAAT GCGCTACGA CTGCGATGAT	9480
GTATGGGATT CCAGTCTTGA TTTTATCTT TGCAGTTTAT GCGCCAGGTG GAGTCGCCCT	9540
ATACTGGACA GTGTCTAATG CTTATCAAGT CTGCAAACC TATTTCTTGA ATAATCCATT	9600
CAAGATTATC GCAGAGCGCG AGGCCGTAGT ACAGGCACAA AAAGATTTGG AAAATAGAAA	9660
AAGAAAAGCC AAGAAAAGG CTCAGAAAAC GAAATAAATA AGGAGGAATC TGGTAGTGGT	9720
AGTATTTACA GGTCAACTG TTGAAGAAGC AATCCAGAAA GGATTGAAAG AATTAGATAT	9780
TCCAAGAATG AAGGCTCATA TCAAAGTCAT TTCTAGGGAG AAAAAAGGCT TTCTTGGTCT	9840
ATTTGGTAAA AAACCAGCCC AAGTGGATAT TGAAGCGATT AGTGAAACGA CTGTTGTCAA	9900
AGCAAATCAA CAGGTAGTAA AAGGCGTTCC GAAAAAATC AATGATTTGA ACGAGCCTGT	9960
GAAGACGGTT ACTGAAGAAA CCGTTGACCT TGGTCATGTG GTTGATGCTA TTAATAAAT	10020
AGAGGAAGAA GGTCAAGGTA TTTCTGATGA AGTCAAGGCT GAAATCTTAA AACATGAAAG	10080
ACATGCCAGC ACTATCTTAG AAGAACTGG TCACATTGAG ATTTTAAATG AACTTCAAA	10140
CGAGGAAGCG ATGAGGGAAG AAGCAGGCGC TGATGACCTT GAACTGAGC AAGACCAAGC	10200
TGAAAGTCAA GAACTAGAAG ACTTGGGCTT GAAAGTTGAA ACGAACTTG ATATTGAACA	10260
AGTAGCTACG GAAGTAATGG CTTATGTTCA AACGATTATT GATGACATGG ATGTTGAGGC	10320
TACACTTTCA AATGATTATA ACCGTCGTAG CATCAATCTA CAAATTGACA CCAACGAACC	10380
AGGTCGTATT ATCGGCTACC ATGGTAAAGT CTTGAAGGCC TTGCAACTGT TGGCTCAAAA	10440
TTATCTTTAC AACCCTATT CCAGAACCTT CTACGTTACA ATCAATGTCA ATGATTATGT	10500
CGAACACCGT GCAGAAGTCT TGCAGACCTA TGCACAAAA TTGGCGACTC GTGTTTGGGA	10560

381

AGAAGGGCGC AGTCATAAAA CAGATCCAAT GTCAAATAGC GAACGCAAGA TTATCCATCG	10620
TATTATTTCA CGTATGGATG GCGTGACTAG TTA CTCTGAA GGTGATGAGC CAAATCGCTA	10680
TGTTGTTGTA GATACAGAAT AAGTAAAATC AGGTTTATCC TGATTTTTTG CTAGTTAGAG	10740
GAGGTTAAAC TGATGTTGAA TAAGATAAGA GACTATTTAG ACTTTGCTGG TTTGCAGTAC	10800
CGTAATCCTG ATAAAGCGGG AGCAGAGCGA GAGAAGATGC TGGCATTCCG CCACAAAGGA	10860
CAAGAGGCCC GAAAGGTTTT TACAGAACTG GCCAAAGCCT TTCAAGCAAG CCATCCAGAA	10920
TGGCAACTCC AACAGACTAG CCAGTGGATG AATCAGGCCC AGCGTTTGAG ACCAGATTTT	10980
TGGGTTTATC TACAGAGAGA CGGACAAGTG ACAGAACCTA TGATGGCCTT ACGTTTGTAT	11040
GGGACATCTA CTGACTTTGG AATTTCTTTG GAAGTCAGTT TCATCGAACC TAAGAAGGAT	11100
GAGCAAACAC TGGGCAAGCA GGCCAAAGTT TTAGACATTC CAACCGTTAA AGGGATTTAT	11160
TATCTAACCT ACTCTAATGG TCAAAGTCAA CGGTGGGAGG CGAATGAAGA AAAGCGTCGT	11220
ACTTTACGCG AGAAGGTGAG AAGTCAAGAA GTTCGAAAAG TTTTAGTGAA GGTAGATGTT	11280
CCTATGACAG AAAATTCGTC TGAAGAAGAA ATCGTAGAAG GCTTATTGAA GTCTTATTCT	11340
AAAATCTTTC CCTATTATCT AGCTACGAGA AAATAAGATA ATTTGTAAAA CATCATAAAT	11400
CATACAGTCC AAGAGTGAAC AGTCCGCTGT GTAATCTTG GTCTTTTTGT TTGCGCTTTC	11460
GCATTATATA ATAACTTAC AAAACAATT CAAAAGGAGA ACAATTATGG AAGTCGTTTC	11520
AAGTGTTCTA AATTGGTTTT CTAGCAATAT TTTGCAGAAAT CCCGCATTTT TCGTAGGTTT	11580
ATTGGTGTTG ATAGGATATG CACTTTTGAA AAAACCTGCC CATGACGTTT TTTCAGGGTT	11640
TGTTAAAGCA ACAGTAGGGT ATATGTTGCT TAACGTGGGT GCTGGTGGTT TGGTTACAAC	11700
CTTTCGTCCA ATCTTAGCAG CTCTTAAC TAATTCCTA CAAATTCCTA ATTTGGTGCAG CGGTTATCGA	11760
CCCTTACTTT GGACTTGCTG CAGCAAACAA CAAAATTGTA GCAGAGTTTC CAGATTTTGT	11820
TGGAAGTACA ACTACAGCTC TATTGATTGG TTTTGGAAATA AATATCTTGC TCGTAGCTCT	11880
TCGAAAGATT ACGAAGGTAA GAACCTCTT TATTACTGGT CACATCATGG TACAACAAGC	11940
TGCAACAGTA TCTCTTATGG TTCTATTCTT AGTACCACAA TTGCGCAATG CTTACGGTAC	12000
AGCAGCGATT GGTATCATCT GTGGACTTTA CTGGGCAGTT AGTTCAAATA TGACTGTTGA	12060
GGCAACTCAA CGCTTGACTG GTGGTGGCGG ATTTGCGATT GGTACCAAC AGCAATTTGC	12120
AATCTGGTTT GTAGATAAAG TAGCAGGACG CTTTGGTAAG AAAGAAGAAA GTTTAGACAA	12180
TCTTAAATTA CCTAAGTTCC TCTCAATCTT CCACGATACA GTTGTGTCAT CTGCTACCTT	12240
GATGCTCGTA TTCTTCGGAG CCATTCTTTT AATCTTGGGT CCAGACATTA TGTCTAATAA	12300

382

AGAAGTCATC ACTTCAGGAA CTCTATTCAA TCCTGCTAAA CAAGATTCTT TTATGTACAT	12360
TATCCAAACA GCCTTTACCT TCTCAGTTTA CTGTTCGTT TTGATGCAAG GTGTCCGAAT	12420
GTTCTGATCT GAGTTGACAA ACGCCTTCCA AGGTATTTCA AACAAATTGT TGCCAGGTTT	12480
ATTTCCAGCG GTTGACGTTG CAGCTTCTTA TGGATTGGT TCTCCAAATG CTGTCTTGTC	12540
AGGATTACC TTTGGTTGA TTGGTCAATT GATTACAATT GTTTGCTCA TCGTCTTTAA	12600
AAATCCGATT CTTATTATTA CAGGATTGT ACCAGTGTTC TTTGACAATG CAGCCATTGC	12660
GGTCTACGCT GATAAACGCG GCGGATGGAA AGCGGCTGTT ATCCTTTCTT TTATATCAGG	12720
TGTCCTTCAA GTTGCTCTAG GAGCTCTTGT TGTGGCCCTT CTCGATTGG CATCTTATGG	12780
TGGCTACCAT GGAAATATCG ACTTTGAATT CCCATGGCTT GGATTGGAT ATATCTTCAA	12840
ATACCTGGT ATTGTTGGT ATGTACTTGT GTGCTCTTC TTGCTTGTTA TTCTCAACT	12900
TCAATTGCC AAAGCAAAAG ATAAAGAGAA ATATTACAAC GGTGAAGTTC AAGAAGAAGC	12960
TTAGTATCTA GAAAAGGAGA AATAAAATGG TTAAAGTATT AGCAGCGTGC GGAAATGGAA	13020
TGGGTTATC AATGGTTATC AAGATGAAGG TTGAAAATGC TCTCCGTAAG CTTAATCAAA	13080
CAGATTTTAC AGTCAATTCA TGCAGTGTG GTGAAGCTAA AGGTTTAGCA GTAGGATATG	13140
ACATCGTAAT CGCTTCTCTT CATTTGATTC AAGAATTGGA AGGGCGAACT AATGGGAAGT	13200
TAATTGGGCT TGATAACTTG ATGGATGATA AAGAAATCAC CGAAAACTC AGTCAAGCAC	13260
TACAGTAAA GGTGGAGGG GGCTGGACAG AAAGTGAAG TTATCGTTTC TGTCCTTCTC	13320
CCTCTTAAA TAAAGGAGGC AGATATGAAT TTAACAACAG CTTTAATTGA CAATGACTCG	13380
ATCCGACTAG GTTTAGAGGC TAACAATTGG AAAGAAGCAG TCAAGGTAGC AGTAGATCCC	13440
TTAATTGAAA GTGGGGCAAT TTGCCAGAG TATTACGATG CTATCATGA ATCGACTGAA	13500
GAGTATGGGC CTTACTATAT CTTGATGCCA GGTATGGCTA TGCCCCACGC TAGACCTGAA	13560
GCAGGTGTGC AAAGTGATGC CTTTTCATTG ATTACCTTAC AAAATCCTGT TGTATTTTCA	13620
GATGGGAAAG AGGTATCTGT TTTGTTGGCA CTAGCAGCAA CAAGTTCAAA AATTACACA	13680
AGTGTAGCCA TTCCACAAAT TATTGCCCTA TTTGAATTAG AAGATTCTAT TGCACGTTTA	13740
CAGGCTTGCC AGACTAAAGA AGATGTCTTG GCTATGATTG AAGAATCTAA GGATAGCCCT	13800
TATCTCGAAG GATTGGATTT GGAAAGTTAG AAAGAGGAAT AAAGAAATGA CAAAAAGAAT	13860
ACCTAATTTA CAAGTTGCAT TAGACCATTC AGACTTGCAA GGAGCGATTA AAGCAGCTGT	13920
TTCTGTTGGT CAGGAAGTAG ATATTATCGA AGCTGGAAC GTTTGCTTGC TTCAAGTTGG	13980
AAGTGAAGT GCTGAAGTCT TGCGTAGCCT TTTCCAGAT AAGATTATTG TGGCAGACAC	14040
AAAATGTGCT GATGCTGGTG GAACAGTTC TAAAAATAAT GCGGTTCTGT GAGCAGACTG	14100

GATGACTTGT ATCTGTTGTG CAACCATCCC TACTATGGAA GCAGCTCTAA AGGCTATCAA	14160
GA CTGAACGA GGAGAACGAG GCGAAATCCA GATCGAGCTT TATGGCGATT GGACTTTTGA	14220
ACAAGCTCAG CTTTGGCTAG ATGCAGGTAT CTCACAAGCT ATTTATCACC AATCTCGTGA	14280
TGCTCTTCTT GCTGGTGAAA CTTGGGGTGA AAAAGACCTT AATAAGGTTA AAAAATCAT	14340
TGACATGGGC TTCCGTGTAT CTGTAACAGG TGGTCTAGAT GTAGATACTC TCAAACCTTT	14400
TGAAGGTATT GATGCTCTTA CCTTTATCGC AGGTCGTGGA ATTACAGAGG CTGTGGATCC	14460
AGCAGGAGCA GCGCGTGCCT TCAAGGATGA AATCAAACGA ATTTGGGGGT AAATCATGGT	14520
ACGTCCAATT GGAATTTATG AAAAGGCAAC CCCAACACAC TGTAATTTGGC TAGAACGTTT	14580
AAATTTTGCC AAGGAGTTAG GCTTTGATTT TGTCGAGATG TCTATTGACG AACGTGACGA	14640
GCGTTTAGCA AGACTTGACT GGAGTAAGGA AGAACGCTTG GAAGTTGTCA AAGCAATCTA	14700
TGAAACTGGT GTTCGTATTC CTTCTATCTG TTTTTCAGGC CATCGTCGCT ACCCATTTGGG	14760
TTCAAAAGAT CCAGTTCTAG AGGAAAAATC TCTAGAACTC ATGAAAAAAT GTATCGAATT	14820
AGCTCAAGAC TTGGGAGTTC GTACGATTCA ATTAGCTGGT TACGATGTTT ACTATGAGGA	14880
AAAGTCACCC CAGACACGCC AACGTTTTAT CAAAAATTTG AGAAAAGCCT GTGACTGGGC	14940
TGAAGAAGCT CAGGTGGTAC TTGCTATTGA AATTATGGAT GATCCTTTCA TCAGTAGCAT	15000
CGAAAAATAT TTGGCTATAG AAAAAGAGAT TGACTCTCCC TTCCTCTTTG TATATCCAGA	15060
TATTGGTAAT GTGCTGTCAT GGCATAATGA TATCTATAGT GAGTTTTATC TTGGTCATCA	15120
TGCCATCGCA GCTCTCCATC TCAAGGATAC TTATGCAGTG ACAGAAAGTT CAAAGGGCCA	15180
GTTCCGAGAT GTACCTTTTCG GGCAAGGTTG TGTCAAATGG GAAGAAGCTT TCGATATTTT	15240
AAAGGAAACC AATTATAATG GACCTTTCCT AATCGAAATG TGGTCTGAAA ATTGTGAAAC	15300
AGTAGAAGAA ACACGCGCAG CCATTCAAGA GCGCAAGCT TTTCTCTATC CACTCATTAA	15360
GAAAGCAGGT TTGATGTAAG ATGAATCAAG TAATCAATGC TATGCGTAAA CGAGTCTGTG	15420
ATGCCAATCA ATCATTGCCA AAACATGGAC TTGTCAAATT TACCTGGGGG AATGTATCTG	15480
AAGTTAATCG CGAACTCGGT GTCATTGTTA TCAAACCATC AGGCGTGGAT TATGACGAAT	15540
TGACACCTGA AAACATGGTA GTGACTGATC TAGATGGTAA GATCCTAGAA GGGGATTTAA	15600
GACCATCTTC CGACCTCCCA ACTCATGTGC AATTATATAA GACTTGGTCA GAAATTGGTA	15660
GTGTGGTTCA CACCCATTCG ACAGAAGCTG TTGGTTGGGC TCAGGCAGGT CGTGATATTC	15720
CTTTCTACGG AACAAACCAT GCAGATTATT TCTACGGTTC AATCCCTTGC GCCCGTAGTT	15780
TGACCAAGGA CGAAGTAGAA GTGGCCTATG AAAAAGATAC TGGCCTGGTT ATCGTAGAAG	15840

384

AGTTTGAACA TCGCGGACTT AACCCGGTTG AAGTACCAGG AATTGTTGTA CGCAATCACG	15900
GTCCATTAC CTGGGGCAAA AATCCAGAGA ATGCTGTTTA TCACTCTGTC GTAGTAGAGG	15960
AAGTATCAAA GATGAATCGC TTTACAGAAC AAATCAATCC AAGAGTTGGA CCTGCTCCCC	16020
AGTACATACT AGAAAAACAC TACCAACGTA AACATGGACC AAATGCTTAT TATGGTCAAA	16080
AGTAAGAACG ATGAAGGAGG AGAAAAAGAT AAATTTAGCT CCTCTTTTTA CATTTGATTT	16140
TTATTGAGAG TAAAGTTGGA GTTGAAGTAA TTTTAAAAGA TTTTMTAGAA ATAGCGCTTG	16200
ATATATATAT GGTAAAATAA AAAGAAATGC TGTGATATCA ATAGATTTGG GGGATTTTMT	16260
AATATGGTAC TGGATAAGGC AAGTTGTGAT TTGCTTCAAT ATTTGATGGA TCAAGAAACG	16320
TCCAAAACGA TTATGGCGAT TTCGAAAGAT TTGAAAGAGT CAAGAAGGAA AATTATTAT	16380
CACATTGACA AAATCAATGC TGCTCTGGGT GACGAGGCGC TTCACATCAT TAGTATTCCA	16440
CGAATTGGTA TTCACTTAAC GGAAGAGCAG AGAGATGCTT GTTGTAAACT ATTATCGGAA	16500
GTAGATTCGT ACGATTATAT CATGAGTGGC CATGAACGTA TGATGATAAT GTTACTATGG	16560
ATAGGTATTT CTAAAGAACG TATTACGATT GAAAAATTGA TAGAGTTAAC AGAGGTATCT	16620
AGGAATACTG TTCTCAATGA TTGAATAGT ATTGCTTATC AACTAACTTT GGAACAATAT	16680
CAGGTGATCT TGCAAGTGAG CAAGTCACAG GGATACAACC TTCATGCCCA CCCTCTTAAT	16740
AAAATTCAGT ATCTTCAATC GCTTCTATAT CATATTTTTA TGGAAGAAAA TGCCACTTTT	16800
GTATCTATTT TAGAAGATAA GATGAAAGAG AGGTTAGATG ATGAGTGTTT GCTTCTGTT	16860
GAAATGAACC AATTTTTTAA GGAACAGGTT CCTTTAGTTG AACAAGATTT AGGGAAGAAA	16920
ATAAACCATC ATGAAATAAC TTTTATGTTG CAGGTTCTAC CTTATTGCT GTTAAGCTGT	16980
CATAATGTTG AACAGTATCA AGAAAGACAT CAGGATATAG AGAAAGAATT TTCTTTGATA	17040
AGAAAAAGAA TAGAGTATCA GGTGTCTAAG AAATTAGGAG AACGGTTGTT TCAAAAGTTT	17100
GAAATTTCTT TGTCAGGACT TGAAGTTTCT CTTGTAGCTG TTCTCCTCCT CTCCTATCGT	17160
AAAGATTGGA ATATTCATGC AGAAAGTGAT GATTTTCGGC AATTAAAACT TGCTTTAGAA	17220
GAATTTATCT GGTATTTTGA ATCACAATC CGAATGGAGA TTGAGAACAA GGATGATTTG	17280
TTACGAAATT TGATGATCCA CTGTAAAGCC TTGTTATTTA GAAAGACTTA CGGTATTTT	17340
TCTAAAAATC CTCTAACAAA ACAAATTCGA TCCAAGTATG GAGAATTATT TTTAGTCACT	17400
AGAAAACTG CGGAAATTTT AGAAGGAGCA TGGTTTATTC GGCTAACAGA CGATGATATT	17460
GCCTATTTGA CGATTCATAT TGGAGGATTT TTAATATA CACCATCATC TCAAAAAAT	17520
ATGAAAAAAG TTTATCTCGT TTGTGATGAA GGTGTTGCGG TTTCGAGACT TTTGCTGAAA	17580
CAATGCAAAC TTTATTTTCC AAATGAGCAA ATGACACTG TATTTACAAC AGAACAAATT	17640

385

AAGAGTGTGG AAGATATTGC ACAAGTTGAT GTAGTGATTA CTACTAATGA TGATTGTGGAT	17700
AGCAGATTTC CGATTTTAAG GGTTAATCCT ATCCTTGAAG CAGAAGATAT TTTGAAAATG	17760
CTAGACTATC TTAAACACAA TATATTTTCGT AATAAGAGCA AAAGTTTCAG TGAAAATCTT	17820
TCTAGTCTTA TTTCGTCTTA TATTGTAGAC AGCAAGTTGG CTAGTAAGTT CCAAGAAGAG	17880
GTTCAAACAC TTATAAATCA AGAAATAGTA GTTCAAGCTT TTTTGAAGr TATTTGAAGG	17940
ACAGTCCAAT GATGAACACA AACCTGTGtK TTTCsTGGTC TTTTtTAGTG TTTTGAAGGG	18000
TGGKATACTA ATCTCAAAGA TAACAATTAT ATCCAAAGGA GGCAACATAT GCCAAACGTC	18060
AAAGAAATTA CAAGAGAGTC ATGGATTTTA GCCACTTTCC CAGAGTGQGG AACATGGTTG	18120
AACGAAGAAA TCGAAGAAGA AGTCGTACCT GAAGGCAACT TTGCCATGTG GTGGCTAGGC	18180
AACTGTGGTA CTTGGATTAA GACACCAGCT GGTGCTAACG TTGTCATGGA CCTTTGGTCA	18240
AACCGTGGAA AATCAACCAA AAAAGTGAAA GATATGGTTC GTGGGCACCA AATGGCAAAT	18300
ATGGCAGGTG TTCGTAAGCT GCAACCAAAC TTGCGTGTTC AGCCAATGGT TATCGATCCA	18360
TTTGCTATCA ACGAACTAGA CTATTACTTA GTTTCACACT TCCACAGTGA TCATATCGAC	18420
CCATACACAG CTGCAGCAAT TCTCAATAAT CCTAAGTTAG AGCATGTTAA GTTGG	18475

(2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7186 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCAGGATTTG GTACCGTTGC AAGTGGTGTG CCTTTCCTCC TAAAGGAAAA TGGAGGAAAA	60
ATCAATCAAT CAGCACATTC AGATATCAAA GTTGCTAAGG TATTGGTCAA GGATGAAGAT	120
GAAAAAATC GCTTGCTTGC AGCAGGGAAT GACTTTAACT TTGTAACCAA TGTGGATGAT	180
ATTTTATCAG ACCAGGATAT TACTATCGTA GTGGAATTGA TGGGGCGTAT TGAGCCTGCT	240
AAAACCTTTA TCACTCGTGC CTTGGAAGCT GGAAAACACG TTGTTACTGC TAACAAGGAC	300
CTTTTAGCTG TCCATGGCGC AGAATTGCTA GAAATCGCTC AAGCTAACAA GGTAGCACTT	360
TACTACGAAG CAGCAGTTGC TGGTGGGATT CCAATTCCTC GTACTTTAGC AAATTCCTTG	420
GCTTCTGATA AAATTACGCG CGTGCTTGA GTAGTCAACG GAACTTCCAA CTTTCATGGTG	480
ACCAAGATGG TGGAAGAAGG CTGGTCTTAC GATGATGCTC TTGCGGAAGC ACAACGTCTA	540

386

GGATTTGCAG AAAGCGATCC GACGAATGAC GTAGATGGGA TTGATGCAGC CTACAAGATG	600
GTTATTTTGA GCCAATTTGC CTTTGGCATG AAGATTGCCT TTGATGATGT AGCCCACAAG	660
GGAATCCGCA ATATCACACC AGAAGACGTA GCTGTAGCTC AAGAGCTTGG TTACGTAGTG	720
AAATTGGTTG GTTCTATTGA GGAAACTTCT TCAGGTATTG CTGCAGAACT GACTCCAACC	780
TTCTTACCTA AAGCGCACCC ACTTGCTAGT GTGAATGGCG TAATGAACGC TGTCTTTGTA	840
GAATCTATCG GTATTGGTGA GTCTATGTAC TACGGACCAG GTGCGGGTCA AAAACCAACT	900
GCAACAAGTG TTGTAGCTGA TATTGTCCGT ATCGTTCGTC GTTTGAATGA TGGTACTATT	960
GGCAAAGACT TCAACGAATA TAGCCGTGAC TTGGTCTTGG CAAATCCTGA AGATGTCAAA	1020
GCAAACACT ACTTCTCAAT CTGGCTCTA GACTCAAAA GTCAGGTCTT GAAGTTGGCT	1080
GAAATCTTCA ATGCTCAAGA TATTTCTTTT AAGCAAATCC TTCAAGATGG CAAAGAGGGT	1140
GACAAGGCGC GTGTCGTAT CATCACACAC AAGATTAATA AAGCCCAGCT TGAAAATGTC	1200
TCAGCTGAAT TGAAGAAGGT TTCAGAATTC GACCTCTTGA ATACCTTCAA GGTGCTAGGA	1260
GAATAAGATG AAGATTATTG TACCTGCAAC CAGTGCCAAT ATCGGGCCAG GTTTTGACTC	1320
GGTCGGTGTA GCTGTAACCA AGTATCTTCA AATTGAGGTC TGCGAAGAAC GAGATGAGTG	1380
GCTGATTGAA CACCAGATTG GCAAATGGAT TCCACATGAC GAGCGTAATC TCTTGCTCAA	1440
AATCGCTTTG CAAATTGTAC CAGACTTGCA ACCAAGACGC TTGAAAATGA CCAGTGATGT	1500
CCCTTTGGCG CGCGGTTTGG GTTCTTCCAG CTCGGTTATC GTTGCTGGGA TTGAACTAGC	1560
CAACCAACTG GGTCAACTCA ACTTATCAGA CCATGAAAA TTGCAGTTAG CGACCAAGAT	1620
TGAAGGGCAT CCTGACAATG TGGCTCCAGC CATTTATGGT AATCTCGTTA TTGCAAGTTC	1680
TGTTGAAGGG CAAGTCTCTG CTATCGTAGC AGACTTTCCA GAGTGTGATT TTCTAGCTTA	1740
CATTCCAAAC TATGAATTAC GTACTCGCGA CAGCCGTAGT GTCTTGCCTA AAAAATGTC	1800
TTATAAGGAA GCTGTTGCTG CAAGTTCTAT CGCCAATGTA GCGGTTGCTG CCTTGTGGC	1860
AGGAGACATG GTGACCGCTG GGCAAGCAAT CGAGGGAGAC CTCTTCCATG AGCGCTATCG	1920
TCAGGACTTG GTAAGAGAAT TTGCGATGAT TAAGCAAGTG ACCAAAGAAA ATGGGGCCTA	1980
TGCAACCTAC CTTTCTGGTG CTGGGCCGAC AGTTATGGTT CTGGCTTCTC ATGACAAGAT	2040
GCCAACAATT AAGGCAGAAT TGGAAAAGCA ACCTTTCAAA GGAAACTGC ATGACTTGAG	2100
AGTTGATACC CAAGGTGTCC GTGTAGAAGC AAAATAAAGA ATAGAAGATA GGATGGGGAA	2160
ACTCTTGACC AGAGGGGTTT ATATCCTTTT TGTGAAAAGA AGTTTATACT CAATGAAAAT	2220
CAAAGAGCAA ACTAGGAAGC TAGCCGCAGG CTGCTCAAAA CAGTGTTTTG AGGTTCGAGA	2280
TAGAACTGAC GAAGTCAGCT CAAGACACTG TTTTGAGGTT GCAGATAGAA CTGACGAAGT	2340

387

CAGTAACCAT	ACTACGGTAA	GGTGACGCTG	ACGTGGTTTG	AAGAGATTTT	CGAAGAGTAT	2400
TAGTTAAAAA	CGTGATAAAG	GAGAAATAAA	GATGGCAGAA	ATTTATCTAG	CAGGTGGTTG	2460
TTTTTGGGGC	CTAGAGGAAT	ATTTTTCACG	CATTTCTGGA	GTGCTAGAAA	CCAGTGTGGG	2520
CTACGCTAAT	GGTCAAGTCG	AAACGACCAA	TTACCAGTTG	CTCAAGGAAA	CAGACCATGC	2580
AGAAACGGTC	CAAGTGATTT	ACGATGAGAA	GGAAGTGTCA	CTCAGAGAGA	TTTTACTTTA	2640
TTATTTCCGA	GTTATCGATC	CTCTATCTAT	CAATCAACAA	GGGAATGACC	GTGGTCGCCA	2700
ATATCGAACT	GGGATTTATT	ATCAGGATGA	AGCAGATTTG	CCAGCTATCT	ACACAGTGGT	2760
GCAGGAGCAG	GAACGCATGC	TGGGTCGAAA	GATTGTCAGTA	GAAGTGGAGC	AATTACGCCA	2820
CTACATTTCTG	GCTGAAGACT	ACCACCAAGA	CTATCTCAGG	AAGAATCCTT	CAGGTTACTG	2880
TCATATCGAT	GTGACCGATG	CTGATAAGCC	ATTGATTGAT	GCAGCAAACT	ATGAAAAGCC	2940
TAGTCAAGAG	GTGTTGAAGG	CCAGTCTATC	TGAAGAGTCT	TATCGTGTCA	CACAAGAAGC	3000
TGCTACAGAG	GCTCCATTTA	CCAATGCCTA	TGACCAAACC	TTTGAAGAGG	GGATTTATGT	3060
AGATATTACG	ACAGGTGAGC	CACTCTTTTT	TGCCAAGGAT	AAGTTTGCTT	CAGGTTGTGG	3120
TTGGCCAAAGT	TTTAGCCGTC	CGATTTCCAA	AGAGTTGATT	CATTATTACA	AGGATCTGAG	3180
CCATGGAATG	GAGCGAATTG	AAGTTCGTTT	TCGTTCAGGC	AGTGCTCACT	TGGGTCATGT	3240
TTTCACAGAT	GGACCGCGGG	AGTTAGGCGG	CCTCCGTTAC	TGTATCAATT	CTGCTTCTTT	3300
ACGCTTTGTG	GCCAAGGATG	AGATGGAAAA	AGCAGGATAT	GGCTATCTAT	TGCCTTACTT	3360
AAACAAATAA	AACAGAGAGT	GGGGCTTCCC	ACTTTCTTCA	TTTCTAGAAT	ATGAATAGAA	3420
GGGATTTATG	AAACACCTAT	TATCTTACTT	CAAACCCTAC	ATCAAGGAAT	CAATTTTAGC	3480
CCCCTTTGTT	AAGCTGTTAG	AAGCTGTTTT	TGAGCTCTTG	GTTCCCATGG	TGATTGCTGG	3540
GATTGTTGAC	CAATCTTTAC	CTCAGGGAGA	TCAAGGTCAT	CTCTGGATGC	AGATTGGCCT	3600
GCTCCTTATC	TTTGCAGTAA	TTGGCGTTTT	AGTGGCCTTG	ATAGCTCAAT	TTTACTCAGC	3660
AAAGGCAGCA	GTAGGTTCTG	CTAAGGAATT	GACAAACGAT	CTTTATCGTC	ATATTTCTTC	3720
CTTGCCCAAG	GACAGCAGAG	ACCGTCTGAC	AACCTCTAGT	TTGGTCACTC	GCTTGACTTC	3780
GGATACCTAC	CAGATTCAGA	CTGGTATCAA	TCAATTCCCTG	CGTCTCTTTT	TACGAGCGCC	3840
CATTATCGTT	TTTGGTGCCA	TTTTTATGGC	TTATCGAATC	TCAGCTGAGT	TGACTTTCTG	3900
GTCTTTAGTC	TTGGTTGCCA	TTTTGACCAT	TGTCATTGTA	GGGTATCTC	GATTGGTCAA	3960
TCCTTTCTAC	AGTAGTCTCA	GAAAGAAAAC	GGACCAACTG	GTTTCAGGAA	CGCGCCAGCA	4020
ATTGCAAGGG	ATGCGGGTTA	TTCTGCTTTT	TGGTCAAGAA	AAACGAGAGT	TACAGATTTT	4080

388

TCAAACCCTT AACCAAGTTT ATGCTAGATT ACAAGAAAAG ACAGGTTTCT GGTCTAGTTT	4140
ATTAACACCT CTGACCTATC TGATTGTCAA TGGAACCTCT CTCGTTATTA TCTGGCAAGG	4200
CTATATTTCA ATTCAAGGAG GAGTGTCTCAG TCAAGGTGCT CTCATTGCTC TTATCAATTA	4260
CCTCTTACAG ATTTTGGTGG AATTGGTCAA GCTAGCCATG TTGATCAATT CCCTCAACCA	4320
GTCCTATATC TCAGTCAAGC GAATCGAGGA AGTCTTTGTT GAGGCTCCAG AGGATATCCA	4380
TTCAGAGTTA GAACAAAAGC AAGCTACCAG AGATAAGGTT TTACAAGTCC AAGAATTGAC	4440
CTTTACCTAT CCTGATGCGG CCCAGCCTTC TCTGAGATAC ATTTCTTTTG ATATGACTCA	4500
AGGACAAATT CTAGGTATCA TCGGGGGAAC TGGTTCTGGT AAATCAAGCT TGGTGCAACT	4560
CTTACTTGA CTTTATCCAG TAGACAAGGG GAACATTGAC CTTTATCAAA ATGGACGTAG	4620
TCCTCTTAAT TTGGAGCAGT GCGGGTCTTG GATTGCCTAT GTACCTCAA AGGTCGAACT	4680
CTTTAAAGGA ACCATTTCGT CCAACTTGAC TCTAGGTTTC AATCAAGAAG TATCTGACCA	4740
GGAACCTCTGG CAGGCCTTGG AGATTGCGCA AGCTAAGGAT TTTGTCAGTG AAAAGGAAGG	4800
ACTCTTGGAT GCTCTAGTTG AGGCAGGGGG GCGAAATTC TCAGGTGGAC AAAACAAAG	4860
ATTGTCTATC GCCCGAGCAG TCTTGCGCCA GGCTCCGTTT CTCATCCTAG ATGATGCAAC	4920
CTCGGCACTG GATACCATTA CAGAGTCCAA GCTCTTGAAA GCTATTAGAG AAAATTTTCC	4980
AAACACGAGC TTAATTTTGA TCTCTCAACG AACCTCAACT TTACAGATGG CGGACCAGAT	5040
TCTCCTCTTG GAAAAAGGTG AGTTGCTAGC TGTGGCAAG CACGATGACT TGATGAAATC	5100
CAGCCAAGTC TATTGTGAAA TCAATGCATC CCAACATGGA AAGGAGGACT AGAATGAAAC	5160
GACAAACTGT AAACCAGACG CTCAAACGTT TAGCCGTAGA TTTAGCAAGC CATCCTTTCC	5220
TCCTTTTCCT AGCCTTTCTA GGAACATTTG CCCAAGTTGG CTTATCAATT TACCTACCTA	5280
TTCTGATTGG GCAGGTCATT GACCAAGTCC TAGTGGCTGG TTCATCACCA GTTTTPTGGC	5340
AGATTTTCTT CCAGATGCTC TTGGTGGTAA TAGGAAATAC TCTGGTACAA TGGGCCAATC	5400
CTCTCCTCTA TAATCGTCTA ATCTTCTCTT ATACCAGAGA TTTACGGGAG CGAATCATCC	5460
ATAAGCTCCA TCGTTTACCG ATTGCCTTTG TAGATAGGCA AGGTAGTGGG GAGATGGTTA	5520
GTCGTGTAAC CACGGACATC GAACAGTTGG CAGCTGGCTT GACCATGATT TTTAACCAAT	5580
TTTTCATTTG TGTTTTGATG ATTTTGGTCA GTATTCTAGC CATGCTCCAA ATTCATCTCC	5640
TCATGACTCT CTTAGTCTTG CTGTTGACGC CACTGTCCAT GGTGATTTCA CGCTTTATTG	5700
CCAAGAAATC CTATCATCTC TTCCAGAAGC AAACAGAGAC GAGGGGAATT CAGACTCAGT	5760
TGATTGAAGA ATCGCTTAGT CAGCAGACTA TAATCCAGTC CTTCAATGCT CAAACAGAAT	5820
TTATCCAAAG ATTGCGTGAG GCTCATGACA ACTACTCAGG CTATTCTCAG TCAGCCATCT	5880

389

TTTATTCCTC AACGGTCAAT CCTTCGACTC GCTTTGTAAA TGCACTCATT TATGCCCTTT	5940
TAGCTGGAGT AGGAGCTTAT CGTATCATGA TGGGTCAGC CTTGACCGTC GGTCTTTAG	6000
TGACTTTTTT GAACTATGTT CAGCAATACA CCAAGCCCTT TAACGATATT TCTTCAGTGC	6060
TAGCTGAGTT GCAAAGTGCT CTGGCTTGCG TAGAGCGTAT CTATGGAGTC TTAGATAGCC	6120
CTGAAGTGGC TGAAACAGGT AAGGAAGTCT TGACGACCAG TGACCAAGTT AAGGGAGCTA	6180
TTTCCTTTAA ACATGTCCTT TTTGGCTACC ATCCTGAAAA AATTTTGATT AAGGACTTGT	6240
CTATCGATAT TCCAGCTGGT AGTAAGGTAG CCATCGTTGG TCCGACAGGT GCTGGAAAAAT	6300
CAACTCTTAT CAATCTCCTT ATGCGTTTTT ATCCCATTAG CTCGGGAGAT ATCTTGCTGG	6360
ATGGGCAATC CATTTATGAT TATACACGAG TATCATTGAG ACAGCAGTTT GGTATGGTGC	6420
TTCAAGAAAC CTGGCTCACA CAAGGGACCA TTCATGATAA TATTCCTTTT GGCAATCCTG	6480
AAGCCAGTCG AGAGCAAGTA ATTGCTGCTG CCAAAGCAGC TAATGCAGAC TTTTTCATCC	6540
AACAGTTGCC ACAGGGATAC GATACCAAGT TGGAAAAATGC TGGAGAATCT CTCTCTGTCG	6600
GCCAAGCTCA GCTCTTGACC ATAGCCCGAG TCTTTCTGGC TATTCCAAAG ATTCTTATCT	6660
TAGACGAGGC AACTTCTTCC ATTGATACAC GGACAGAAGT GCTGGTACAG GATGCCTTTG	6720
CAAAACTCAT GAAGGGCCGC ACAAGTTTCA TCATTGCTCA CCGTTTGTC ACCATTCAGG	6780
ATGCGGATTT AATTCTTGTC TTAGTAGATG GTGATATTGT TGAATATGGT AACCATCAAG	6840
AACTCATGGA TAGAAAGGGT AAGTATTACC AAATGCAAAA AGCTGCGGCT TTTAGTTCTG	6900
AATAAGCCAT TCTCTTTTGA AAGTTTATGG ACGAAAAAAG TTGCCTTCGA GTGACTTTTT	6960
TGTTACAATA GCTAGAAAAA TTGTTCACTG TAATACTCAA TGAAAATCAA AGAGCAAAC	7020
AGGAAGCTAG CCGTAGGTTG CTCAAAGCAC AGCTTTGAGG TTGTAGATAA GACTGACGAA	7080
GTCAGTTCAA AACACTGTTT TGAGGTTGCA GATAGAACTG ACGAAGTCAG CTCAAAACAC	7140
TGTTTTGAGG TTGCAGATAG AACTGACGAA GTCAGCTCAA AACAGG	7186

(2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14273 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CTGAAAATTC TAAAAAATTT ATAAGTAAGG AATTAATTAG TTATTTTGTG GATAAAGTTT	60
---	----

390

ATGATGAAAT ATTTGTTGAA GAGGTAGTTC CGCACGTTTT TCTGCCATAT GAATCTGACT	120
TACTTCTTAT TTTACCAGCT ACGGCAAATG TGATTGGCAA AATTGCTAAT GGTATTGCTG	180
ATGATTTAGT TACAGCAACT GTTTTAACT TTAATAAAAA AATAATTTTT TGTCCCAATA	240
TGAACTCTAC TATGTGGGAC AATCACATAG TTCAAAGAAA TGTATCAATT CTAAAGGAGT	300
TGGGACATAT ATTTTTATTT GAGTCTAAAA AACATATGA GGTAGGATTG CGTAAAGCAA	360
TAGATTCAAC ATGTTCAATG TTACAACCAC AGTCGTTAGT AAAAGAACTT ATCAAATTAG	420
AAAATATTGT CCTTGAAGAG GGACATTAAT AACTACTGAG AATATTAAATG AGGGGAAAAA	480
ATGGAAAAAT CATCAATCGA TGTAGATATG CTGTTGGAAG AATTGACACA AGAAGCAATG	540
GTCGTTGTTG CTGTTGATAA GGACTGTTAA TTTAAACTTA TGGCAATATA TGAAAGGTTA	600
CTGGATGTTT TAAATTATGC AGGCAGTAGC CTTTTATTAT ATACAAATGG ATAAAGTAAG	660
GATAATACAA TGATTAATAA AAAAATACAA CAAGTTGTTT TGGAAATCATT ACAGAATTTT	720
TTGAATGGGA ACTTCATTTT GCCTTGTTGA GTCTATGATT TTGGCTTGCT GGAAACTGTA	780
CTTGATGAAT TTAATAATCA AATTCTGTGA ACATTCAATT ACCAACTTTT TTATGCCGTT	840
AAAGCAAATT CAAATGAGAA GATACTTGAA TTCTTAGTAG ATAAAAATTGA TGGAGTTGAT	900
GTGGCGTCAT TATCTGAATT AGATGTGGCT AAAAAATTTT TCCCACCAAC TCAAATTTCT	960
GTTAATGGTC CCGCATTTTC TTATGAACT TTATATAATC TGATTAAAAA ACAATATAAA	1020
GTTGATATTA ACTTTTTGGA ACATCTTCAA CAATTTTCCC CAAAAGAATC TGTTGGAATA	1080
AGAGTAACGG AGCCAGATGA ACTTAATAAT CGTATGAGTC GATTGGAAT AAATATTTGC	1140
AGTGATAATT GGACTAGTAA TTTACAAAAT CCTTTAATTA CACGACTGCA TTTTCATTTT	1200
GGAGAAAAAG ATGATAAATT TATTGTTAAG TTAGATAAAA TATTATTTAA GTTACAAGAA	1260
ATTAATAAAC TTAGAGAGGT TAGAGAAATA AATCTTGGAG GCGGTTTTAT GAAATTTTTT	1320
ATGGAAAAATC GTTTGAAAGA ATTTTCTCTA TCACTTATGG AAATCTATAA AAAGTACGAT	1380
ATTGATAGTA CTGTGACTAC AATAATAGAA CCAGGTAGTG CAATTACTTC ATTTTCTGCC	1440
TATATGATTA CTAGCCCAGT TAATGTTAGT GAGGTGAATG AGCAGCAGGT TATCACGTTA	1500
GACACATCAA TATACACCAA TACATTATGG TTTGTTCCGC ATATTATTAC AACGTTAAAT	1560
TCAAGTAGTA AAGAGCGTTA TAGTACTATT CTCTATGGTA ATACCTGTTA TGAACATGAC	1620
AAGTATAAAA TGAAAGTTTC GCTTCCAAGG TTAACATAA ATAGCAGTAT AGTGTTTTTT	1680
CCTGTAGGAG CTATATATAA AAGCAATCAT TCAAATTTAC ATCGTAATGA TTTTATGCGG	1740
GAGGTATATT TGTGGACAAA AAACCTTGACA TATTAGATAA AGTTAAGGAA TATTTAGGAA	1800
ATAAACTAC TCAAATCTG GATAATCAAT ATAAAGAATT TTTGAAACTT AATGATATAA	1860

391

GGCGAGCGTT TGGTATTTCA GAAAAAGTAT TAAACAATTC TTTTAATTTT ACGAGTAAAG	1920
AATTTAATGA TTTAATTAAT AACGAAATTT ATTTATTCGA ATATGCATGT AGAATTAGAG	1980
AGGAATGGAG AAAAAATGC TTTAATCATT CTTATCGTTT TCTATGCTCA CCTATAATTA	2040
CAGATGATTT TCTTAACACG AAGACATTGA GAAGTAGCCA AATTGAATAT AAATATGAGC	2100
GATATTTATC GAAAAGTTCG ATAGGCGATA GAGCGGTTGA TGGCTTTGTT TCCTTCAATA	2160
CTTTAACAGC TAATGGTATG TCTGCTATTA AACTATGTCT TGAGATATTA AACTCTATTT	2220
TCTTCAAGAA GAAGATTGAT TTATTATATT CAACCGGATA TTATGAAACA AGATTTTAT	2280
TAAATAATCT TGCTAAATCA GGTATTAGTT GCTATGAGGT AAGTAATTGT GAATTGGATA	2340
AAGATAAATT TTATAATGTA TTCATGATGG AACCCAATCG AGCCGATTTA ACATTACAAA	2400
AAACTGATTT CAAGATAGTA GAATATTTTG TTAAGTATAA AAATAATTCA ATAAAAGTCG	2460
TTATTTTAGA TATTTTCATAT CAAGGTTCTA ATTTTAAATT AGTAGAATTT TTAGAGAAAT	2520
TTAAATTTGC GAATGTAATT ATTTTGTGG TACGATCTTT GATAAAATTA GATCAAATGG	2580
GATTAGAATT GACAAATGGG GGAATAATAG AAGTGTATAT TCCTAATCAT TTGAGAAAGT	2640
TGAAAAATTT TATTGAAGAG GAATTCATAA AATTTAGAAA TTCTCACGGA GCTAATCTAA	2700
GCCTCTATGA ATACTGTTTG CTTGATAATT CTTTAACTTT AAAAAATGAT TGAAGTATT	2760
CTGATTTAGT TATGAAATTT ACGAGTAATT TTTATGCTGA TATAAAAGAC TTGTTCATGG	2820
AAAATCTGA TATTGAAATC ATCCATGAAG AGGGAGTACC TTTTGTATTT TTAGATTTAA	2880
TAGGTGAAGG TAAAAAGAA TATGAAATGT TTTTCAATG GTTAAACTTC TTTTACAAAC	2940
AGCTTGGAAT CACATGTAT GCTAGAAATA GTTTTGGGTT TCGGAATCTA ACAGTAGAGT	3000
ATTTTGGAAT TATTGGGACA GAAAGATATA TATTTAAGAT TTGTCCAGGT GTTTATAAAG	3060
GGTTAAGTTA TTATTTGATG AAATTTTAT TAAATCTTT TTCAAATGAA TATTTAAAAA	3120
CTACTGATGA GGTTAATAGA TGAAAAATTT GATAAAGTTG CTAATAATTA GATTGATTGT	3180
TAACTTAGCA GACAGTGTAT TTTATATAGT AGCATTGTGG CACGTTAGCA ATAATTATTC	3240
TTGAGCATG TTCTTAGGAA TATTTATTGC AGTAAATTAT CTACCGGATT TGTTACTAAT	3300
CTTTTTTGGG CCAGTTATTG ACAGAGTAAA TCCGCAAAAA ATTCTTATAA TATCAATTTT	3360
GGTTCAATTA GCAGTGGCTG TAATATTTTT ATTATTATTA AACCAGATAT CATTTTGGGT	3420
GATAATGAGT CTAGTGTATA TTTCAGTAAT GGCTAGCTCC ATAAGTTACG TGATAGAAGA	3480
TGTGTTGATT CCTCAAGTGG TAGAATATGA TAAGATTGTA TTGCAAATT CTCTTTTATG	3540
TATTTCTGAT AAAGTATTAG ATTCTATTTT TAATTCATTC GCATCATTTT TACAGGTGGC	3600

392

AGTAGGATTT ATTTTATTGG TTAAGATAGA TATAGGCATA TTTTACTTGG CTCTATTTAT	3660
ATTGTTGTTG TTAAATTTA GAACTAGCAA TGCGAATATA GAAAACCTCT CTTTCAAATA	3720
TTACAAGAGA GAAGTGTGCG AAGGTACAAA GTTTATTTTA AATAATAAAT TATTATTTAA	3780
AACCAGTATT TCTTTAACGC TTATAAACTT TTTTATTCA TTTCAGACAG TAGTTGTACC	3840
GATTTTCTCT ATTCGATATT TTGATGGTCC GATTTTATAT GGTATTTTAT TAACTATTGC	3900
TGGTTTGGGT GGTATATTGG GAAATATGCT AGCGCCAATC GTAATAAAAT ATTTAAAATC	3960
GAATCAAATT GTTGGTGAT TTCTTTTAT GAACGGCTCA AGTTGGTTAG TAGCAATTGT	4020
TATAAAAGAC TATACTTTAT CACTTATTTT ATTTTTCGTT TGTTTTATGT CTAAAGGAGT	4080
CTTCAATATT ATTTTAAATT COTTGTACCA ACAAATACCT CCACATCAAC TTCTTGGTAG	4140
GGTAAATACT ACCATTGATT CTATTATTC TTTTGAATG CCAATTGGTA GTTTAGTTGC	4200
AGGAACGCTT ATTGATTTGA ATATTGAATT AGTGTAAAT GCTATTAGCA TACCTTATTT	4260
TTGTTTCTCT TATATTTTAT ATACGGATAA TGGATTGAAA GAATTAGTA TATATTAGAA	4320
ATGTTTATGT TCATTCAAAA GCATAATGAC TATAACTGAA AAAGAAAAGT GATATCTTTA	4380
AGGTTGTTCT TCTTGGTGGT GAGATTCGTG AGACAACCCA AGCTTTTGTG GGAAAGATTA	4440
CCAATGCTTT GATGGATAGG ATGTACTTTA GCAAGATGTT TTTAGTGGTA ACGGTATCGT	4500
GGATGGACGT GTAATAACCT CTTCTTCGA GGAGTATTTT ACTAAAAAAC TAGCCTTGGA	4560
GCGTCCCCCA GAAACGGACT TACTCATTGA CTCTCAAAG ATTTGGGGAG AAGATTTTGC	4620
TTTCATCTGTT CCTTGAAAA AGTCACAGCA GTCATCACAG ACGATAGTAC TGAACAAAAC	4680
TATGAAGAGT TAGAAATTTA TACGCAGGTG ATGTATATAA GGATCTGGAA ATAGATAAGA	4740
AGTTGATTAG TATTGACCTA GGTGGTACAA ATATTAAGAT TACTGTTCTT TCAAATGACG	4800
GTGAGATTGA AACTTTGTGG AGTATTACAA CAGATACAAG TGAGAAAGGT TCTCAAATTA	4860
TATCGGACAT CATCAGTTCT ATTAATAAATA AATTGACCGA ACGGAATATT CCTGATAGCG	4920
ACCTTCTTGG AATCGGTATG GGAAGTTGCT CATCATACTT TCCTTGTAAT TCATAGGGGC	4980
TATAAACTCT CCGTCTACTT GTCCTGCAAC AATTGAAGTC TGCTCAAAAC GCCGTCCGCT	5040
AATCTTTTCA TAGACTTTCT CCCTTTTAGG AGCCTAGCTT TCTAGTTTGT TCTTTGATTT	5100
TTATTGAGTA TACCACTATT TTAATCCCTC TGGCAAGGGA CTTTGTCTAT GTGGAGGGAT	5160
TGGGCTCCTA TGTGGTGGAG CTTTCTGTT CTTTCTGAAA TATGGTATAA TAGCACTAAT	5220
CAATTTCTAG GAAAATAGAT ACAGAAAGGG GCTGAAAGAT GTCTCATATT ATTGAATTGC	5280
CAGAGATGCT GGCAAACCAA ATCGCGGCTG GAGAGGTCAT TGAACGTCCT GCCAGTGTGG	5340
TCAAAGAGTT GGTAGAAAAT GCCATTGACG CGGGCTCTAG TCAGATTATC ATTGAGATTG	5400

393

AGGAAGCTGG TCTCAAGAAG GTTCAAATCA CGGATAACGG TCATGGAATT GCCCAGCATG	5460
AGGTGGAGTT GGCCCTGCGT CGCCATGCGA CCAGTAAGAT AAAAAATCAA GCAGATCTCT	5520
TTCGGATTCG GACGCTTGGT TTTCGTGGTG AAGCCTTGCC TTCTATGCG TCTGTTAGTG	5580
TCTTGACTCT GTTAACGGCG GTGGATGGTG CTAGTCATGG AACCAAGTTA GTCGCGCGTG	5640
GGGGTGAAGT TGAGGAAGTC ATCCCAGCGA CTAGTCCTGT GGAACCAAG GTTGTGTGG	5700
AGGATCTCTT TTTCAACACG CTGCCCCTG TCAAGTATAT GAAGAGCCAG CAAGCGGAGT	5760
TGTCTCATAT CATTGATATT GTCAACGTC TGGGCTTGGC CCATCCTGAG ATTTCTTTTA	5820
GCTTGATTAG TGATGGCAAG GAAATGACGC GGACAGCAGG GACTGGTCAA TTGCGCCAAG	5880
CAATCGCAGG GATTTACGGT TTGGTCAGTG CCAAGAAGAT GATTGAAAT GAGAACTCTG	5940
ACCTAGATTT CGAAATTTCA GGTTTGTGT CCTTGCCTGA GTTGACTCGG GCTAACCGCA	6000
ATTATATCAG CCTCTTCATC AATGGCCGTT ATATTAAGAA CTTCTGCTC AATCGTGCTA	6060
TTTTGGATGG TTTTGAAGC AAGCTTATGG TTGGACGTT TCCACTGGCT GTCATTACAC	6120
TCCATATCGA CCCTTATCTA GCGGATGTCA ATGTGCATCC AACTAAGCAA GAGGTGCGGA	6180
TTTCCAAGGA AAAAGAACTG ATGACTCTGG TTTCAGAAGC TATTGCAAAT AGTCTCAAGG	6240
AACAAACCTT GATTCAGAT GCCTTGAAA ATCTTGCCAA ATCGACCGTG CGCAATCGTG	6300
AGAAGGTGGA GCAAACTATT CTCCCACTCA AAGAAAATAC GCTCTACTAT GAGAAAACTG	6360
AGCCGTCAAG ACCTAGTCAA ACTGAAGTAG CTGATTATCA GGTAGAATTG ACTGATGAAG	6420
GGCAGGATTT GACCCTGTTT GCCAAGGAAA CCTTGGACCG ATTGACCAAG CCAGCAAAAC	6480
TGCATTTTGC AGAGAGAAAG CCTGCTAACT ACGACCAGCT AGACCATCCA GAGTTAGATC	6540
TTGCTAGCAT CGATAAGGCT TATGACAAAC TGGAGCGAGA AGAAGCATCC AGCTTCCCAG	6600
AGTTGGAGTT TTTCGGACAA ATGCACGGGA CTTATCTCTT TGCCCAAGGG CGAGATGGAC	6660
TTTACATCAT AGATCAGCAC GCTGCTCAGG AACGGGTCAA GTACGAGGAG TACCGTGAAA	6720
GCATTGGCAA TGTTGACCAA AGCCAGCAGC AACTCCTAGT GCCCTATATC TTGGAATTTT	6780
CTGCGGATGA TGCCCTGCGT CTCAAGGAAA GAATGCCTCT CTTAGAGGAA GTGGGCGTCT	6840
TTCTAGCAGA GTACGGAGAA AATCAATTTA TTCTACGTGA ACATCCTATT TGGATGGCAG	6900
AAGAAGAGAT TGAATCAGGC ATCTATGAGA TGTGCGACAT GTCCTTTTGG ACCAAGGAAG	6960
TTTCTATCAA GAAATACCGA GCAGAGCTGG CTATCATGAT GTCTTGCAAG CGATCTATCA	7020
AGGCCAATCA TCGTATGAT GATCATTCAG CTAGACAACT CCTCTATCAG CTTTCTCAAT	7080
GTGACAATCC CTATACTGT CCTCACGGAC GTCCTGTTTT GGTGCATTTT ACCAAGTCGG	7140

394

ATATGGAAAA GATGTTCCGA CGTATTCAGG AAAATCACAC CAGTCTCCGT GAGTTGGGGA	7200
AATATTAAAA GTATAAAAA GTCTGGGAAA AATTTTCAAA ATCAAAAAAA CGCATAAAAAT	7260
CAGGTGTTCA AAAACCTTGA TTTTATGCGT TTTATCATGG AAATAGTTAC TTCATTTTTT	7320
CCTAATTCTT TTCGAAACTC TTTTAAACG ACGTCAGTTT TATCAGTAAT CTCAAAACAG	7380
TGTTTTGAGC TAATTTTGCC AGTTTGTCT GTAACATCGA AGTTGTGTTT TACCACTCTG	7440
CGACTGGTTT CCTAGTTTGC TCTATGATTT TCACAGAGCA TTAAATTGCG ATTTTGCCAA	7500
GTTTCTTTAT TCGTCTAAAA GTAGAGTCTG TTCTATGCGT CTAATGTACG AATCAGGTTG	7560
ACCATTTCAA TAGCTCCTTG TGCACACTCA GAACCTTAT TTCCTGCTTT AGTACCAGCT	7620
CGTTCATGCG CTGTGTTCAAT TGTATCTGTC GTTAGCACAC CAAACATAAC AGGAATTTGCG	7680
CTATTTAAAC TGATTGGGC GATTCCCTTA GATACCTCGC TACATACATA ATCATAATGA	7740
CTTGATTCC CTCTAATGAC AGCTCCCAAG CAGATAATTG CATCATATTT TTTACTTTTT	7800
GCCATTTTTG ATGCAATCAG TGGTATTCA AAAGCTCCTG GAACCCAGGC TACCTCTATA	7860
TCFTTCTCGT TTACATCTC TCTTTTGAGA TTATCTAGTG CTCCAGATAA TAATTTTGAA	7920
GTTATAAATT CATTAAATCT CGCTACAACA ATACCTATTT TAATATGTT TGCTACTAAA	7980
TTACCTTCAT AAGTGTTTAT TTATTTTTCC TCCATATTTA AAATGTGACC CATTCGATTT	8040
TTCTTTGTTT CTAAATAAAA ACTATCGTAA GGATTGGCTT CTATTTGAT TGATATTCTA	8100
CTGGAAATGG TAATCCATA TTTTCTAAC TGTTCACCT TGTCAGGATT ATTTGTCAGT	8160
AAATGAAGTG ACTGAAGTCC CAGATCTTTA AGCATTTTGT CTCCAATATG ATATCTCTTT	8220
AAATCACCTT CAAAGCCTAA TGCAAGATTG GCATCAAGCG TATCCATGCC TTGATCTTGT	8280
AAATGATAGG CTTTAAATTT ATTGATAAGT CCAATTCCTC GTCCCTCCTG TCGCAAGTAA	8340
AGTAAGACAC CCGAACCATT CTCAACAATC ATTTTCATAG CCTTATCGAA TTGCTGTCCA	8400
CAATCGCAAC GTAAAGAGCC TAAAACATCT CCTGTTAAAC ATTCGGAGTG GACCCGACAT	8460
AATACATTGG CTTTATCCTC TATATTTCCC ATAATAAGAG CAAGATGATG TTCCCATTTT	8520
AGTTTATCTA TATAGCTAAT TGCTTTGAAA TTACCGTATC TAGTAGGCAT ATTGACAGTT	8580
GAAACTCGTT CTACCAGCTG ATCATATACT TTTCTATATT CTTGTAATTC TTTGATGGTA	8640
ATTAGTGGA TGTGTGTTT TTTGAGAAC TGAATTAAAT CATCTGTTCT CATCATTTTG	8700
CCATCATGAT TCATTATTTT ACAACATAGG CCACACTCTT TTAGTCCAGC TAATTTTAAT	8760
AAATCAACAG TTGCTTCTGT GTGTCCATTT CTTTCTAGGA CACCACCTTT TTTTGCAATT	8820
AAAGGAAACA TGTGCTCTGG CCTGCGAAAA TCAGAGGGTG TTATATCTTC AGCTACACAC	8880
ATACGTGCGG TCAGTCTCTT TTCTCGGCA GAAATACCTG TGGTCTTTT TTTATAATCA	8940

395

ATTGAAACTG TAAAAGCAGT CTTATGATTA TCTGTATTCT TTTCAACCAT AGGTGAAAGC	9000
ATTAATTGAT TAGCTAAACT TTCGCTCATA GGCATACAAA TTAATCCTTT GGCATAAGTA	9060
GCCATAAAAT TAACATTTTC TGTGTAGCT GCTTGTGCAG AACAAATTAA GTCTCCTTCA	9120
TTTTCTCTAT CCTTGTCTGC TATAACAAGA ACAAGTCGTC CCTTCTGCAA TGCTTCTAAT	9180
GCTTCTTGTA TTTTTCGATA TTCCATTGAC TGATTATCCT TTCTGCTAAA ATCCATTTTG	9240
ATATAATAGT TCCTTAGATA TTTCTGATTT TGGAGAGTTA TCCATCAGTT TTTGCACATA	9300
TTTACCTAAG ATATCATTTT CAAGATTTAC TGTACTCCCG ACTTGTTTAC TCTTAAGAAT	9360
GGTTTGTTC AAGGTATGAG GGATAACAGA TACTGAAAAG TTTACTTTGG AGACTTTAGC	9420
GACAGTCAGA CTAATGCCGT CAATGTAAAT AGATCCTTTT TCAACTATTA AATCTAAAAT	9480
TTCTTTTGT GTGTGATTT GATACCATA AGCATTATCA TCTTTTFTA TTGACGAGAT	9540
TTTTCTGTA CCATCAATGT GTCCTGTAAC GACGTGACCC CCAAGTCGAC CGTTGACAGA	9600
TAAGGCTCTT TCTAGATTCA CCTCACTTCC ATGTTTAAAT AGAGTAAGAG CTGTTGCACT	9660
CCATGTTTCA TTCATTACAT CAACTGTAAA GGATTGATGA TTGAAATGAG TAACTGTAAG	9720
ACAGATACCA TTTACTGCTA TACTATCGCC TAAATGGATA TCCGTTAATA TTTTGTAGGC	9780
TTAATTGAT AGTTTACAAT TACGAGAGTC TTCTGTATT CTTTCAACTT TTCCGATTTT	9840
TTCAATTATT CCTGTGAACA TGGATAAATC ACTTCACTTT CTATGAGATA GTCATTTTCT	9900
ATTTGAGAAA ATGCATAAGG TTTCAATCTA ATAGCGTCAT TTGGCAAAGA AATACCTTCA	9960
CCTCCGACAG GAAACTTGGC ACTACCTCCA AAAACTTTTG GTGCAATATA TATTTTCAGC	10020
TCATCAACAA TTTGTTGTTC CAAAGCACTC CAATTCATTA GACTGCCCCC TTCTAGAACT	10080
AGGCTATCAA TCTGCATGTT TCCTAGATGT TGCATTAAAC TCGATAAGTC TATATGATTG	10140
CCTTTTTTCT TTATGGAAAG TATTTACAG CCATGATTTT GATATAGCTT CATTTTATTT	10200
TTGTCTTCTAG AGGAAGTGGC AATGTAAGTT TTAATATCAT TTGCTGTTT TACGATTTTA	10260
GAGGTAAGAG GAGTTCGTAA ATGTGTATCG CATATGATAC GGATAGGATT TTCCCTTCC	10320
TCCAATCTAC ATGTCAGCAA AGGATCGTCT TGAATAACAG TATTGACTCC CACCATAATT	10380
GCACTAACAT GGTGTCGTAA CTGATGCACA TGCTTTCTTG CTTCTTCTTC AGTAATCCAT	10440
TTGGATTGAT TTGTTTATAGT GGCTATTTTT CCATCCATTG ACATTGCATA TTTCATAAAA	10500
ACATAGGGTA CATGCTGGGT AATATACTTT CTAAACTTTT TTATTAAGTT AAGACACTCA	10560
TTTTCTAAAA TTCCAACAGT AACTGAAGA TTATTTTCTT CAAGTATCTT TACTCCTTTT	10620
CCAGATACAA TAGGATTACA GTCTAGGCTT CCAATGACTA CTCTGTAAAT ACCACTATCG	10680

396

ATTATAGCAT CTATACAGGG AGGTGTTTTT CCGAAGTGAC AACAGGGTTC AAGTGTTACA	10740
TAAAGCGTCG CTCCGACAGG GGATTCTCTA CAGTTTTTAA GAGCATTTCT CTCAGCATGT	10800
GGGCCACCAA AAAACTCATG ATAACCTTGT CCGATAATGT GATTATCTTT TACAATAACT	10860
GCGCCGACCA TAGGATTGGG ATTGACGTAA CCAGCCCCTT TTTGTGCCAG TTTTATTGCT	10920
AATTTTCATAT ATTTTGAATC GCTCATCTCG CTACCTCCAA AAAATATAC CTGAATAGG	10980
GGACTACTCA AGGCATACAA AAGAAAACCT ATGCGATTAA CAAAAATGCT CTGAAATGAC	11040
AAGTAATCAT TTCAGAGCAC GCAAAAAGCA CAAATATACT TTTATCTTCT TTCATCCAGA	11100
CTATACTGTC GGCTTTGGAA TTTCACCAA TCATGCCCTT CGGCTCGTGG GCTATACCAC	11160
CGGTAGGGAA TTTCACCTG CCCTGAAGAT AGTTATTCAA TTACAGATGA TTATAGTACT	11220
TAATTTTGAA TATGTCAACA GATAAATACC GATTGTTTTT GATATACTGT ATTTGTGATA	11280
ATCGATTCTC GCTCCTCGGA TAAAGAAAAT ATGATATACT AGATAAACGA AATAAGAGAG	11340
AAGGAATACT ATGTACGCAT ATTTAAAAGG AATCATTACC AAAATTACTG CCAAATACAT	11400
TGTTCTTGAA ACCAATGGTA TTGGTTATAT CCTGCATGTG GCCAATCCTT ATGCCTATTC	11460
AGGTCAAGTT AATCAGGAGG CTCAGATTTA TGTGCATCAG GTTGTGCGTG AGGACGCCCA	11520
TTTGCTTTAT GGATTTGCT CAGAGGATGA GAAAAGCTC TTTCTTAGTC TGATTTGCGT	11580
CTCTGGGAT GGTCTGTAT CAGCTCTGC TATTATCGCT GCTGATGACA ATGCTGGCTT	11640
GGTTCAAGCC ATTGAAACCA AGAATCATCAC CTACTTGACC AAGTTCCCTA AAATTGGCAA	11700
GAAAACAGCC CAGCAGATGG TGCTGGACTT GGAAGGCAAG GTAGTAGTTG CAGGAGATGA	11760
CCTTCCTGCC AAGGTCGCAG TGCAAGCAAG TGCTGAAAAC CAAGAATTGG AAGAAGCTAT	11820
GGAAGCCATG TTGGCTCTGG GCTACAAGGC AACAGAGCTC AAGAAAATCA AGAAATTCTT	11880
TGAAGGAACG ACAGATACAG CTGAGAACTA TATCAAGTCG GCCCTTAAAA TGTGGTCAA	11940
ATAGGAGCAG AGAATGACAA AACGTTGTTC GTGGGTCAAG ATGACCAACC CGCTCTACAT	12000
CGCCTATCAT GATGAGGAGT GGGGCCAGCC CCTCCATGAT GACCAAGTAT TGTTTGAGTT	12060
GTTGTGTATG GAAACCTATC AGGCAGGCCT GTCTTGGGAA ACGGTACTCA ACAAACGCCA	12120
AGCTTTCCGA GAAGTCTTTC ATAGCTATCA AATCACTCA GTCGCAGAGA TGACTGACAC	12180
TGAATTGGAA GCCATGCTGG AGAATCCAGC TATCATTCGA AATAGAGCCA AGCTTTTTCG	12240
TACACGCGCT AACGCCAAG CCTTCTACA GTTACAGGCA GAGTACGGCT CTTTGTATGC	12300
CTATCTTTGG TCTTTTGTG AGGGGAAAAC TGTCGTTAAC GATGTTCTCG ATTATCGCCA	12360
AGCGCCAGCT AAAACACCCT TATCTGAGAA ATTAGCCAAA GATCTCAAAA AACGAGGCTT	12420
CAAGTTCACA GGCCAGTCG CCGTATTGTC TTTCTACAG GCTGCAGGC TAGTTGATGA	12480

397

CCACGAGAAT GATTGTGAGT GGAAAGGTCT TAAATGATGT CTAACAAAAA TAAGGAAATT	12540
CTGATTTTTC CGATTCTCTA TACAGTCCTC TTTATGTTTG ATGGCGTTAA ATTGCTGGCT	12600
TCTTTAATGC CATCTGCCAT TGCAAAATTAT CTTGTTTATG TAGTTTITAGC TCTATATGGC	12660
TCCTTCTTGT TCAAGGATAG ATTGATCCAA CAATGGAAGG AGATTAGAAA GACTAAAAGA	12720
AAATTCTTCT TTGGAGTCTT AACAGGATGG CTCTTCTCA TTCTGATGAC TGTGTCTTT	12780
GAATTTGTAT CAGAGATGTT GAAGCAGTTT GTGGGACTAG ATGGACAAGG TCTAAATCAG	12840
TCTAATATTC AAAGTACCTT TCAAGAACAA CCACTACTGA TAGCTGTTTT TGCTTGTGTC	12900
ATTGGACCTC TGGTAGAAGA ATTATTTTTC CGTCAGGTCT TATTGCATTA CTTGCAGGAA	12960
CGGTTGTCAG GTTTACTAAG CATTATCTCG GTAGGACTTG TTTTGTCTCT GACTCATATG	13020
CACAGTTTGG CTCTATCAGA GTGGATTGGT GCAGTTGGTT ACTTAGGTGG AGGCCTTGCC	13080
TTTTCTATTA TTTATGTGAA AGAAAAAGAG AATATCTACT ATCCCCTACT TGTTCACATG	13140
TTAAGCAACA GCCTCTCCTT AATCATTTTA GCTATCAGTA TAGTAAAATG AAATGAGAAC	13200
AGGACAAATC GATTCTAAC AATGTTTTAG AAGTAGAGGT GTACTATTCT AGTTTCAATA	13260
TACTGTAATA TGTGATGAAA ATGCCAGTAA TGATACCGAG AAAAAAGCTG AGAAACTPTT	13320
CCCAGCTTTA TTTGTTATAG TCAAAGAGAA TGACTTGTTT CTGTGCATCT ACATGAGCAT	13380
GGACCCCAAA GGGTACAATT GCTCTTGGAG TTGCGTGGCC GACATTCAGA TTATAGACAA	13440
TCGGGATATT GCTGTCAATG ATATCCAATA GTGCCTCTTT ATAGTCGTCA TGGAAAGTTT	13500
CATCCATAGG TTTTCCGACC AAGAGTCCAT TGATGACCGC GAATATGCCA GTGTCTTTA	13560
AAGTTAGCAA CATCTTTTTC AAGTCTTCTG GCTTAGGCTT TTCTTCGCTT GTTTCGAGCA	13620
AGAGGATTTT CCCTTCCCAG TCTGACAAGT CAGGGAAAAG TTTGTATTTT TGGCAGAGTT	13680
CCGTGCTATC TGCGTATCGA GAGTTGTCAA AGATATCGTA GAGGGATTTC AGGCAACCAC	13740
CGAGGATTTT CCCCTCGAAC TGGGCACTTC CTTGCAACAA GTCAAAACCT GTATTTGTAT	13800
GACTGACACG AGGTGTTCCC AGGGCCGTGG GACTAAAATC AGTTCGTTCC TCATACCAAA	13860
CGTCACTAGG GCGGATTCTT GAAATCTTTC CCGTCTCAAT CAATTCTTTA AAGTAGTGAA	13920
GGCTATAGGC TAGCATTTCT TTGTCTAATT CACAAATGTC TGCTAAAAAG GATTGACCAT	13980
AAAAAGTCTT GATTCCCTAAT TTATGCAACA TGAGGTGGTT CATGGTTGTA TCCGAGAAGC	14040
CAAGAAAAAT TTTTGTCTTG ATAACCTTTT GGAGTTGGTC ATTTTCAAAA AGATAAGGTA	14100
GCAAGCGATA GGTATCGTCT CCACCGATGG CACATAGGAT CATGTCGATG CTATCATCAG	14160
AAAAGGCATG AATCAATCC TCTGCACGAG CTTCAGGATG GTCCTTGATA AAGTCTAATC	14220

398

CTTTTAACGA ATGGGGCAAA AAGATGGGAT TGGTCCCAGA TCCTTGAGAC GTT

14273

(2) INFORMATION FOR SEQ ID NO: 41:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GTGAAGTGGC GCAAAAGGTG CAAGTGATGA GCTCAGGTTT TTTAGCTCTT GACATTGCCCC	60
TTGGCTCAGG TGGTTATCCT AAGGGACGTA TCATCGAAAT CTATGGCCCC GAGTCATCTG	120
GTAAGACAAC GGTGCCCCTT CATGCAGTTG CACAAGCGCA AAAAGAAGGT GGGATTGCTG	180
CCTTTATCGA TCGGAACAT GCCCTTGATC CAGCTTATGC TCGGCCCTT GGTGTCAATA	240
TTGACGAATT GCTCTTGTCT CAACCAGACT CAGGAGAGCA AGGTCTTGAG ATTGCGGGAA	300
AATTGATTGA CTCAGGTGCA GTTGATCTTG TCGTAGTCGA CTCAGTTGCT GCCCTTGTTT	360
CTCGTGCGGA AATTGATGGA GATATCGGAG ATAGCCATGT TGGTTTGCAG GCTCGTATGA	420
TGAGCCAGGC CATGCGTAAA CTTGGCGCCT CTATCAATAA AACCAAAACA ATTGCCATTT	480
TTATCAACCA ATTGCGTGAA AAAGTTGGAG TGATGTTTGG AAATCCAGAA ACAACACCGG	540
GCGGACGTGC TTTGAAATTC TATGCTTCAG TCCGCTTGA TGTTCGTGGT AATACACAAA	600
TTAAGGGAAC TGGTGACCAA AAAGAAACCA ATGTCGGTAA AGAACTAAG ATTAAGGTTG	660
TAAAAATAA GGTAGCTCCA CCGTTTAAGG AAGCCGTAGT TGAAATTATG TACGGAGAAG	720
GAATTTCTAA GACTGGTGAG CTTTGAAGA TTGCAAGCGA TTTGGATATT ATCAAAAAAG	780
CAGGGGCTTG GTATTCTTAC AAAGATGAAA AAATTGGGCA AGGTTCTGAG AATGCTAAGA	840
AATACTTGGC AGAGCACCCA GAAATCTTTG ATGAAATTGA TAAGCAAGTC CGTTCTAAAT	900
TTGGCTTGAT TGATGGAGAA GAAGTTTCAG AACAAAGATC TGAAAACAAA AAAGATGAGC	960
CAAAGAAAGA AGAAGCAGTG AATGAAGAAG TTCCGCTTGA CTTAGGCGAT GAACTTGAAA	1020
TCGAAATTGA AGAATAAGCT GTTAAAGCAG TGGAGAAATC CGCTACTTTT TCGATTTTGT	1080
ATTCAAGTTT TTAGATTATA TATAGTAGCT TGAAATAAGA TATGAACAAC TCTATTAGGA	1140
AAGTCAAATT AATTTCTAGA AATGTTTGTAG CAGCTACAGC GTACTATTCC AAACCTCAACC	1200
AACTATAATA GATCGAAACT AGAATAGTAC ATATCTACTT CTAAAACATT GTTAAAAATC	1260
GATTTGACTT TCCTTATTTT ATTCGGCTAT ATATAGTTTG CTGTTTCTTG TCGCTCCTCT	1320
GGAAAGCTGA TATAATAGCT TTATGAATAA AAAACGAACA GTGGACCTGA TACATGGTCC	1380

399

GATTCCTCCC	TCGCTCTTAA	GCTTCACCTT	TCCAATTTTG	CTATCAAATA	TTTTTCAACA	1440
GCTCTATAAC	ACTGCTGATG	TCTTGATTGT	TGGACGATTT	CTTGGTCAAG	AATCCTTGGC	1500
TGCAGTAGGA	GCGACGACAG	CGATTTTGA	CCTGATTGTA	GGTTTACAC	TGCGTGTGG	1560
CAATGGCATG	GGGATTGTCA	TTGCTCGTGA	TTATGGGGCT	CGGAATTTC	CTAAAATCAA	1620
GGAAGCAGTA	GCAGCCACCT	GGATTTTAGG	TGCTCTTTTG	AGCATTCTAG	TTATGTTGCT	1680
GGGCTTTCTT	GGCTTGATC	CTCTCTTGCA	ATACTTAGAT	ACTCCTGCAG	AAATTCCTCC	1740
TCAATCTTAT	CAATATATTT	CTATGATTGT	GACCTGTGTA	GGTGTGAGCT	TTGCTTATAA	1800
TCTTTTGA	GGCTTGTTGC	GGTCTATTGG	TGACAGTCTA	GCAGCCCTGG	GATTTCTGAT	1860
TTTCTCTGCC	TTGGTTAATG	TGGTCTGGA	TCTCTATTTT	ATTACGCAAT	TGCATCTGGG	1920
AGTTCAATCC	GCAGGACTTG	CTACCATTAT	TTGCAAGGT	TTATCAGCGG	TTCTCTGCTT	1980
TTATTATATT	CGTAAAAGTG	TGCCAGAACT	CTTGCCACAG	TTTAAACATT	TCAAATGGGA	2040
CAAAAGCTTG	TACGCGGATC	TCTTGAGCA	AGGTTTGGCT	ATGGGCTTGA	TGAGTTCAAT	2100
TGTATCTATC	GGCAGTGTGA	TTTTACAGTT	TTCTGTTAAT	ACATTTGGTG	CAGTGATTAT	2160
TAGTGCCAG	ACGGCAGCTC	GACGCATTAT	GACCTTTGCC	CTTCTTCTTA	TGACCCGTAT	2220
TTCTGCATCA	ATGACGACCT	TTGCTTCTCA	GAATCTAGGA	GCTAAGCGAC	CTGACCGTAT	2280
TGTTCAAGGT	CTTCGAATCG	GCAGTCGTTT	AAGTATATCC	TGGGCAGTTT	TTGTTTGTAT	2340
TTTCTCTTTT	TTTGCCAGTC	CAGCTTTGGT	TTCTTCTTTG	GCTAGTTCGA	CAGATGGTTA	2400
CTTGATAGAA	AATGGAAGTC	TCTATCTGCA	AATCAGTTCA	ACCTTTTATC	CCATTTTGAG	2460
CCTCTTGTG	ATTTATCGCA	ATTGCTTGCA	GGGCTTGGGG	CAAAAGATCC	TTCTCTAGT	2520
TTCTAGCTTT	ATTGAACTAA	TCGAAAAAAT	CGTTTGTGTG	GTTTTGATTA	TTCTTGGGC	2580
AGGATATAAG	GGTGTATCC	TTTGTGAACC	TCTTATCTGG	GTGCCATGA	CAGTTCAACT	2640
GTACTTCTCA	TTATCCGTC	ATCCCTTGAT	AAAAGAAGGC	AAGGCAATCT	TGGCAACCAA	2700
AGTGCAATCC	TAGTTGGATT	TACTGAATAA	AATCCATTTT	CTCTAGTGAA	AATCGAAAAA	2760
ACTTGTGTTT	TCTTCTTTAG	TTTGGTGTG	AAAATAGTTT	AACAGACTTT	TGACTTCTTT	2820
TATATGATAT	AATAAAGTAT	AGTATTTATG	AAAAGGACAT	ATAGAGACTG	TAAAAATATA	2880
CTTTTGAAAA	TCTTTTGTAGT	CTGGGGTGT	ATTGTAGATA	GAATGCAGAC	CTTGTCAGTC	2940
CTATTTACAG	TGTCAAATA	GTGCGTTTGT	AAGTTCTATC	TACAAGCCTA	ATCGTGAATA	3000
AGATTGTCTT	CTTTGTAAAG	TAGAAATAAA	GGAGTTTCTG	GTTCTGGATT	GTAATAAATG	3060
AGTTGTTTTA	ATTGATAAGG	AGTAGAATAT	GGAAATTAAT	GTGAGTAAAT	TAAGAACAGA	3120

400

TTTGCCTCAA	GTCGGCGTGC	AACCATATAG	GCAAGTACAC	GCACACTCAA	CTGGGAATCC	3180
GCATTCAACC	GTACAGAATG	AAGCGGATTA	TCCTGGCGG	AAAGACCCAG	AATTAGGTTT	3240
TTTCTCGCAC	ATTGTTGGGA	ACGGTTGCAT	CATGCAGGTA	GGACCTGTTC	ATAATGGTGC	3300
CTGGGACGTT	GGGGGCGGTT	GGAATGCTGA	GACCTATGCA	GCGGTTGAAC	TGATTGAAAG	3360
CCATTCAACC	AAAGAAGAGT	TCATGACGGA	CTACCGCCTT	TATATCGAAC	TCTTACGCAA	3420
TCTAGCAGAT	GAAGCAGGTT	TGCCGAAAAC	GCTTGATACA	GGGAGTTTAG	CTGGAATTAA	3480
AACGCACGAG	TATTGCACGA	ATAACCAACC	AAACAACCAC	TCAGACCACG	TTGACCCTTA	3540
TCCATATCTT	GCTAAATGGG	GCATTAGCCG	TGAGCAGTTT	AAGCATGATA	TTGAGAACGG	3600
CTTGACGATT	GAAACAGGCT	GGCAGAAGAA	TGACACTGGC	TACTGGTACG	TACATTTCAGA	3660
CGGCTCTTAT	CCAAAAGACA	AGTTTGAGAA	AATCAATGGC	ACTTGGTACT	ACTTTGACAG	3720
TTCAGGCTAT	ATGCTTGACG	ACCGCTGGAG	GAAGCACACA	GACGGCAACT	GGTACTGGTT	3780
CGACAACTCA	GGCGAAATGG	CTACAGGCTG	GAAGAAAATC	GCTGATAAGT	GGTACTATTT	3840
CAACGAAGAA	GGTGCCATGA	AGACAGGCTG	GGTCAAGTAC	AAGGACACTT	GGTACTACTT	3900
AGACGCTAAA	GAAGGCGCCA	TGGTATCAAA	TGCCTTTATC	CAGTCAGCGG	ACGGAACAGG	3960
CTGGTACTAC	CTCAAACCAG	ACGGAACACT	GGCAGACAAG	CCAGAATTCA	CAGTAGAGCC	4020
AGATGGCTTG	ATTACAGTAA	AATAATAATG	GAATGTCTTT	CAAATCAGAA	CAGCGCATAT	4080
TATTAGGTCT	TGAAAAAGCT	TAATAGTATG	CGTTTTCTTG	TGGAGATATT	TCCTTCAATT	4140
TTGCTACTAT	ATTAAACAAA	AATCAAAAAG	CAAACTAGAA	AGTTATGCTC	AAATAAAATC	4200
TAAATTTGAC	AATGTAAACC	GAGTCGGATA	GCTTTAAGTA	CTGTTTTGAG	GTTGAAGATA	4260
CGATTTTGA	TAGGAACCTA	TCAATTTTAG	ATTTTTAAGC	AGCATCAATA	AATTGCTTCC	4320
TTGTTTTGTC	ATAATTTTTT	TATTTAAAAA	ATTATGACma	GAGTGTGCTA	TTCTTTTAT	4380
GAGAGGTGTA	TGAATATGAT	AAATGTATGT	GATAAATGTA	TGTGATGTTG	GAAAAAGAAT	4440
AAAAGAACTT	AGAATATCTT	CAAATCTTAC	TCAAGATAAG	ATTGCTGAGT	ATTTGTCTTT	4500
GAATCAAAGC	ATGATTGCCA	AAATGGAAAA	AGGTGAAAGG	AATATCACGA	ATGGATTTAA	4560
GTAATAAAGC	TTCAAATCTT	AGAAAAAAGT	TGGGAGCTGA	TGGTGAATCG	CCGATAGATA	4620
TTTTTAAATT	GGTACAAAAG	ATAGAAAATT	TGACGCTGGT	ATTTTATGGA	CTCGGAAAGA	4680
ATATTAGCGG	AGTCTGTTAT	AAAGGAACTC	AGTTCAGTCT	CATTGCAGTC	AATTCAGACA	4740
TGCCATTAGG	AAGGTAAAGA	TTTTCTTTAG	CACATGGACT	GTATCATCTT	TATTATGATG	4800
AGGTGAAGAA	GAGTTCAGTC	AGTCTTATCT	TGATTGGTGA	AGGAGATGAA	ACTGAAAGAA	4860
AAGCGGATCA	GTTTGCTTCT	TATTTTTTAA	TTTTCCCATC	TTCATGTAT	AGGATGGTTG	4920

401

AGGAAATCAG AGAAAATGCC AATAGAATC ATCTTGAAGT AGAAGATATT ATAAAATTGG	4980
GTCAGTTTTA TGGTATCAGT CATAAAGCTA TGTATATAG ATTGAGGAAT GATGGATACC	5040
TTGATGCAGA AGAAATTAAA AATATGGATA TTAGTGTTAT AGAGACAGCT TCAAGATTAG	5100
GCTATGATAC AAGTTTATAT CGTCCTTTGT CAGAAAGTAA AAAAGAAATG GCATTAGGAT	5160
AATATATTAA TTCAACTGAA CAACTTTTAG AAAATAACAG AATTTCCGAA GGGAAGTATG	5220
AGGAACTGTT ACTAGATGCT TTCAGATATG ATATTGTATA TGGGCTAGAT GAAGAGGGGG	5280
GAGTTGTCGT TTGACTAGTC GTGTATTTAT TGATGCAGAT TGTATTTTATG TATTTTATG	5340
GGTTGGCACT GAACATCTTT TAGAAAAGCT CTATTTGGGT AAAATTGTTA TTCCACAAGA	5400
GGTGTATGAT GAAATCAATA TACCTACAAT TCCCCATTTA AAATCTAGGA TAGATCAGTT	5460
GGTAGCTAAG GGTTCACTG AGATTGTGAG CATAGACATT GGAAGTGAAG AATACGCATT	5520
ATATAGAGAT TTAACAAGAA ATCATGATAG TAACAAGATT ATTGGTAAGG GAGAAGGGGC	5580
ATCTATTTCC TTAGCGAAAA AGCATAATGG GATATTAGGA AGTAATAACC TAAGAGATGT	5640
TAAATCATAT GTAGAAGAAT TTTCTTTAGA ATATATGACA ACAGGAGATA TACTGATTGA	5700
AGCGTTTAAA GCGTAATTTA TTACTGAATA AGAGGGCAAT CATATCTGGA ATAATATGCT	5760
TAAAAAGAGA AGGAAAATTG GTGCAAAATC ATTTTCAGAC TATCTTCGTG GAAGTATTCA	5820
TCAAAAATAGA CAAAAATAAA TTTGGATAAA TCGAACTCAC TATTCAGGAG GCATATGAGC	5880
AATTCGAAAA AGAAAAGTGT CAAATTGAGC CTATAGGAGT AGAAGTGAAA TAGTAAGTCC	5940
TGCATAGTGG ATGAGAGAAA AGTTCTCCTT GAAGTTTTC TGAAGTATCA GTCGCATGTC	6000
AAACGATATG TAGGGTAATG TGAGAGGGGA TAGCGAGTAG TTTTGGTTA TTTTATCAAA	6060
AAACTTATAT TTTATTATAC CGAATGATAA AATATAATAA AAATGATAGA ATAAGGAAAA	6120
AACATGAATG TCAAAAAGAT AATGTCAATT TTTCAATCCT TTTATGTTGA TGTCAGTATT	6180
GAGGAACTGA CTTTGACTTT ACCAATCAGT TTTGTAAAAA GGTTTGAGTA TACTCAAATG	6240
ACTTTTCATA AGGAATCATT TTTATTGATT AAAGAAAAGA GAAGGGGGAG TTTGAGTTCA	6300
TTTGTTACTC AGGCTCGCAC TATGGGTGAA AAAGCCAATA TGGATGTTGT TTTGGTGT	6360
TCGAAGTTAT CAGACAGTGA AAAAAAGCAA TTACTTCAAG CTAGAGTTCC GTTTGTAGAC	6420
TTTAAGGGAA ACCTCTTCTT CCCTCCATFG GGACTAGTAC TCAATGCGAA TGATACTGAA	6480
GTCCCTAAGG AATTACACC TAGCGAACAA TTAACGTGGA TTGCCTTTTT ATTGACAAAA	6540
GGTCAAAAAG TAGTAGATGT TGATTTGCTT TCACAAGTCA CTGGACTTCC AAACCAACA	6600
ATTTATAGGT GTTTGAGGAC TTTTAAAGCT TTATATTGGT TAAACAAGCA AAATAAGCTT	6660

402

TACACATATA CGGTGTCAAA GAAAGAATTA TTCTTAAAT CCGTGTCATG TTTATTTAAT	6720
CCCATCAAAA AACGGATTTT ATTGCCAGAT GCGGATATAA AGCAGATAAA ATCTGTTTCT	6780
AACCTTCTAT ATGGTGGTGC TTATGCTTTG TCGCATTCOA CTTTTTTAGC TGAAACGGAT	6840
GAAAATATTA GCTATGTCAT ATGGCAGAGA AAATTCAATC AGTTATCCTT GCCACTTTCT	6900
CAGCATGTTT TAAATGAAA GATGCTAGAG ATATGGAAAT ATCGTCCTTT TGTATCTGAG	6960
TTTTGGAAATG ATTTTAAAA TAATCATGAT AAACAATTTG TAGATCCGAT TTCTCTTTAT	7020
TTGACCTTAA AAGATGATGA TGACCCACGT ATAGAGGAAG AGAGTGAAGC ACTAGAAAA	7080
ATGATATTAC AGTATCTGGG AGAAGATGAT GCCAGCTAAT ACGAAAGTTA TTTTTCAAGA	7140
AATGTTTTCG GATTTTCAGA ACTATTATGT TCTGATTGGG GGAAGTCTA CCTCTATCGT	7200
ATTGGATTTC CAAGGATTTA AAAGTCGCAC AACAAAAGAT TATGATATGG TCATCATTGA	7260
TGAAGTAAAA AATAAGGAAT TTTATACTAC CTGAATCAT TTTTAGAAT TGGGAGAGTA	7320
TCAAGGAAGT CAGAAAGATG AGAAAGCGCA GCTTTTTTCGA TTTACAACAA CTAATCCTGA	7380
GTTTCCTTCT ATGATTGAAC TATTTAGTAT CTTACCAGAA TATCCATTAA AGAAGGACGG	7440
TCGAGAAAT CCCTTACATT TTGACCAAGA TGCTAGTTTA TCAGCCTTAT TATTGGATGA	7500
AGATTATTAT AATATATTGG TGCATGAAAA AGAAACCATT CAGGGGTATT CGGTATTGAG	7560
TAATTGTGGT TTATACTCTT CGAAAATCTC TTCAAACCCAC GTCAGCTTCC ATCTACAACC	7620
TCAAAACAGT GTTTTGAGCA GCCTGCAGCT AGCTTCCTAG TTTGCTCTTT GATTTTCATT	7680
GAGTATTAAT TATTTTAAAG GCTAAAGCTT GGCTGGATAT GAGGGAGCGC TCTGCCACAG	7740
GTGCTCAAGG TTTAAGTAAG TCCATTAAAA AGCATTTGAA TGACCTTACC CGTTTGACAG	7800
CTTCCTTGCT AGGAGATGAA AAGTTATCGG CTATAACATC AAGTAGTCCG GTAAAAGCAG	7860
ACATGCACCG CTTTGTGATA GAATTAGAGC CTGTGAAGTC AACTATTCTT CAAAATAATG	7920
ACATTTCAAT GGATCAAAAT GAAATTTTGG AAATTTCTGAA AAATTTTCTC GATGGTTAAA	7980
ATAATTGTAG CGAGATGGCT ATATTGAATT CGTCTATATC TGGAACTAG AAAAACTTC	8040
AATTCAGGA GAAAATGAAG TCAATCTTCC CACAATCAA CGTATAGTAT CAAGGTTTTT	8100
CAAGACCTGA TATTATGCGT TTTTGTCTT TCAAAACTTT TTGCCCAGTC TTCGTTTTTA	8160
TCCTCTAGTC ACTTGATTG TTTGAGTGG TTTTGTAGTA TAGTAGAATG AAACGAGAAC	8220
AGGACAAAT GATCAGGACA GTCAAATCGA TTTCTAACAA TGTTTTAGAA GCAGAAGTGT	8280
ACTATTCTAG TTCAATCTA CTATAGTTAA ATCTGCGGTC AAGTCTACTG GTGAATCTAT	8340
GATTGTAATA CTCTTCCAAA ATCTCATCAA CCACGTCAGT CTGCGCTGTC AGTCTGTATC	8400
TTACTGACCA AGCTAGTGAT GGATTTAGAA TAGGTGATTT GGAGCGTCCT ATTAGCTAGG	8460

403

AAATGCTGCT CATAGTCCTT TGCTGAGGCT AGGGTGTTC AACATTCAAC ACTCAACTGG	8520
TTGATCTAGT TGATAGGAAG GGAGTTACTA TAAAATACTC AGGCTTCCAT CATATTTTTT	8580
GAAACGATG TGTAATCAAA ATGTACCAAT ATTGTAGTAT TGGTACAGAA GATGTTGTGA	8640
ATGGATAAAT ATATCATAAC TGCTATCTCA AAAAGATTTC ATATGTCTGT GCATATATAA	8700
TAGACTTCCT GCAAACTAG AATCCTAGTT CATGATTGAT AATACCAGCA ATCAAATTCA	8760
TTCGTAATCC AAAGCGTTTA CGATGATTTC GATAGGTTGT TGAAAACATT TTAACGTTT	8820
CTACTTTGGC AAAGATGTTT TCAACCTTGC TTCTCTCCTT AGATAGCGCA TGGTTATAGG	8880
CTTTATCTTC AGCTGTTAGC GGCTTGAGTT TGCTGGATTT ACGTGGAGTT TGTGCTTGAG	8940
GACATATCTT CATGAGCCCT TGATAACCAC TGTCAGCCAA GATTTTACCA GCTTGTCCGA	9000
TATTTCTGCA ACTCATTTTG ACAAACCTCA TATCATGACA ATAGTTCACA GTGATATCCA	9060
AAGAAACAAT TCTCCCTTGA CTTGTGACAA TCGCTTGAGC CTTCATAGCG TGAAATTTCT	9120
TTTTACCAGA ATCATTCGCT AATTCTTTTT TTAGGGCGAT TGATTTTAC TTCCGTCGCA	9180
TCAATCATTA CCGTGTCTC AGAACTAAGA GGAGTTCTTG AAATCGTAAC ACCACTTTGA	9240
ACAAGAGTTA CTTCAACCCA TTGGCTCCGA CGGATTAAGT TGCTTTCGTG AATACCAAAA	9300
TCAGCCGCAA TTTCTTCATA AGTGC GGAT TCTAGGCTTA ATTTAGGTTT TCGTCCACCT	9360
TTTGC GTT TAAGTTGATA AGCTGTTTTT AATACAGCTA ACATCTCTTT AAAAGTCGTG	9420
CGCTGAACAC CAACAAGACG CTTAAATCGT GTATCAGTTA ATTGTTTACT TGCTTCATAA	9480
TTTCGCAGGG AGTCTATTGA CTCTTGGA GGTGTCAATG TTTTTTTCAT CTATCCCAG	9540
AATTATTTT CCGCCATTTG TATTTGCAA TGCTGAGTAG GTTTCCAGA AAGACTCTGG	9600
AAGATTGTTT TTAGCTTTTT TGTATTCTAA ATCAACCCCT TCAAATTTTA AGTCCATATT	9660
TTTCCTTTAC ATCTGTTTTT TGTGGTCTG GTATTTGTTT AAGTTGAGTG ATAATATAGC	9720
GAATTGAATT TCGAGAGTTT TTAATCAGTT AATTCTTTTT TTAACCCACT TTAATTGCTT	9780
TTTAAACACG GGTAAAAAA GAAATTAAAG TGGGTTAATT TTTCTTGA	9828

(2) INFORMATION FOR SEQ ID NO: 42:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3369 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

404

CCGCGAAAGA TATTTTGA CAAGAGTTTG GACGTGAGGT CCGTGGCTAT AATAAAGTAG	60
AAGTTGACGA GTTTTGTAGAC GATGTCATCA AGGACTATGA AACCTATGCT GCCTTGGTCA	120
AGTCACTTCG TCAGGAAATT GCGGATTTGA AGGAAGAATT AACTCGTAAA CCGAAACCTT	180
CACCAGTTCA AGCAGAACCC CTTGAAGCGG CAATTACAAG TTCTATGACG AATTTTGATA	240
TTTGAACG CCTGAATAGA TTGGAAAAAG AAGTTTGTG TAAACAAATT TTAGATAACT	300
CAGATTTTGA AGTAGTTATT TGAGATGTGC AATTTTGGGA TAATCGCGTG AGGAGAATTG	360
TTTCTCATGA GGAAAGTCCA TGCTAGCACA GGCTGTGATG CCGTAGTGT TTGTGCTAGG	420
CGAAACCATA AGCCTAGGGA CGAGAAATCG TTACGGCAGT TGAAATGGCT AAGTCCTTGG	480
ATAGGCCAGA GTAGGCTTGA AAGTGCCACA GTGACGGAGT CTTTCTGGAA ACAGAGAGAG	540
TGGAACGCGG TAAACCCCTC AAGCTAGCAA CCCAAATTTT GGTCGGGGCA TGGAGTACGC	600
GGAAACGAAC GTAGTATTCT GACTGCTATC AGCTAGAGCT GTTAGTGGTA GACAGATGAT	660
TATCGAAGGA AGTGGTCCTA GTCACTTCTG GAACAAAACA TGGCTTATAG AAAATTGCAT	720
ATAGGTTGGG GCTGAGAAAT TTTCTCAACC TCATTTTGA AAGTGGACAT ATAGAAAGGT	780
CTTGCAAGAC TGTAACATGA AAAAAGAATT TAATTTAATT GCAACTGTGG CAGCAGGGCT	840
TGAGGCTGTC GTTGGTCGTG AAGTGCGAGA GTTGGGCTAC GATTGTCAGG TTGAAAATGG	900
ACGTGTTTCGT TTTCAAGGAG ACGTGAGAGC TATTATCGAA ACCAACCTTT GGCTTCGGGC	960
AGCAGATCGT ATCAAAATTA TCGTAGGAAC GTTCCCAGCT AAGACTTTTG AAGAGCTATT	1020
TCAGGGAGTT TTCGCTTTGG ATTGGGAAAA TTATTTACCA CTGGAGCTC GGTTCCTGAT	1080
TTCAAAAGCT AAATGTGTTA AGTCCAAACT TCACAATGAG CCCAGTGTTC AGGCTATTTC	1140
TAAGAAAGCT GTTGTCAAGA AATTGCAGAA AACTATGCT CGCCCAGAAG GGGTCTCTCT	1200
GATGGAGAAT GGCCAGAGT TTAAGATTGA GGTCTCTATT CTCAAAGATG TGGCAACTGT	1260
CATGATTGAT ACGACCGGT CTAGCCTCTT TAAACGTGGT TATCGTACCG AAAAAGGTGG	1320
CGTCTCTATC AAGGAAAATA TGGCAGCAGC CATTTTACAA CTTTCTAACT GGTATCCAGA	1380
CAAGCCTTTG ATTGATCCGA CCTGTGGTTC GGGGACTTTC TGTATTGAGG CAGTTATGAT	1440
TGCTAGAAAG ATGGCGCCAG GTCTTCGTCG CTCTTTTGCA TTTGAGGAAT GGAAGTGGAT	1500
CAGCGATCGC TTGATTCAAG AAGTGCGCAC AGAAGCGGCT AAAAAAGTAG ACCGTGAGCT	1560
TGAGCTGGAT ATCATGGGCT GTGATATTGA TGCTCGCATG GTGGAAATTG CTAAGGCCAA	1620
TGCTCAGGTA GCTGGTGTG CAGGAGACAT TACTTTTAAG CAGATGCGCG TGCAGGATTT	1680
ACGTTCCGAT AAAATCAATG GAGTAATCAT TTCCAATCCG CCTTATGGTG AACGTTTGTG	1740
AGATGATGCA GGGGTGACCA AGCTCTATGC TGAGATGGGG CAAGTATTTG CACCGCTGAA	1800

405

AACTTGAGC	AAATTTATCC	TGACTAGTGA	TGAAGCTTTT	GAAAGCAAGT	ATGGTAGCCA	1860
AGCAGATAAG	AAGCGTAAGT	TATACAACGG	AACCTTGAAA	GTGGATCTAT	ATCAATATTT	1920
TGGTCAGCGT	GTCAAACGGC	AAGAGGTAAA	ATAGAAAGGG	ATACTCATGA	GTAAAAAAAG	1980
ACGAAATCGT	CATAAAAAAG	AAGGTCAAGA	ACCGCAATTT	GATTTTGATG	AAGCAAAAGA	2040
GCTAACAGTT	GGTCAAGCTA	TTCGTAAAAA	TGAAGAAGTG	GAATCAGGAG	TCTTGCCTGA	2100
GGATTCCATT	TTGGACAAGT	ATGTTAAGCA	ACACAGAGAT	GAAATTGAGG	CGGATAAGTT	2160
TGCGACTCGT	CAATACAAAA	AAGAGGAGTT	CGTTGAAACT	CAGAGTCTGG	ATGATTTAAT	2220
TCAAGAGATG	CGTGAGGCTG	TAGAGAAGTC	AGAAGCTTCT	TCGGAGGAAG	TTCCATCTTC	2280
TGAAGACATC	TTACTACCCT	TGCCTCTGGA	CGATGAGGAG	CAAGGCTTGG	ATCCTCTATT	2340
GCTAGATGAT	GAAAATCCAA	CAGAAATGAC	TGAAGAAGTG	GAAGAGGAGC	AAAACCTTTC	2400
TCGTCTGGAT	CAAGAGGACT	CAGAAAAGAA	AAGTAAAAAA	GGCTTTATTT	TGACCGTTTT	2460
GGCGCTTGTA	TCAGTAATTA	TTTGTGTCAG	TGCTTATTAT	GTCTACCGTC	AAGTGGCTCG	2520
TTCGACTAAG	GAAATTGAAA	CTTCTCAATC	AACTACAGCC	AATCAATCGG	ATGTGGATGA	2580
TTTTAATACA	CTTTATGACG	CCTTTTACAC	AGATAGCAAT	AAAACGGCTT	TGAAAAATAG	2640
CCAGTTTGAT	AAACTGAGTC	AACTCAAGAC	TTTACTTGAT	AAGCTGGAAG	GTAGTCGTGA	2700
ACATACGCTT	GCCAAATCTA	AATATGATAG	TCTAGCAACG	CAAATCAAGG	CTATTCAAGA	2760
TGTCAATGCT	CAATTTGAGA	AACCAGCTAT	TGTGGATGGT	GTGTGGGATA	CCAATGCCAA	2820
AGCCAAATCG	GATGCTAAAT	TTACGGATAT	TAAAACCTGA	AATACGGAGC	TTGATAAAGT	2880
GCTAGATAAG	GCTATCAGTC	TTGGTAAGAG	CCAGCAAACA	AGTACTTCTA	GCTCAAGTTC	2940
AAGTCAAAC	AGCAGCTCAA	GTTCAAGTCA	AGCAAGTTCA	AATACGACTA	GTGAGCCAAA	3000
ACCAAGTAGT	TCAAATGAGA	CTAGAAGTAG	TCGCAGTGAA	GTCAATATGG	GTCTCTCGAG	3060
TGCAGGGGTT	GCTGTTCAAA	GAAGTGCCAG	TCGTGTTGCC	TATAATCAGT	CTGCTATTGA	3120
TGATAGTAAT	AACTCTGCCT	GGGATTTTGC	GGATGGTGTC	TTGGAACAAA	TTCTAGCGAC	3180
TTACAGTTCA	CGTGGCTATA	TCACTGGAGA	CCAATATATC	CTTGAACGTG	TCAATATCGT	3240
TAACGGCAAT	GGTTATTACA	ACCTCTACAA	GCCAGATGGA	ACCTATCTCT	TTACCCCTTAA	3300
CTGTAAGACA	GGCTACTTTG	TCGGAAATGG	CGCTGGTCAT	GCGGATGACT	TAGATTACTA	3360
AGCAGTCGG						3369

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9713 base pairs

406

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

AAGTTTACAA TTTAAATGAA TTAACAATTT TCCCACTAA AAGCACTCCA GTTACCGCAA	60
CGTTTGTA CT GAATGTA CT AATCGCAT TC CATCAACT TC ATCTGTTTCG TCAACTTGAA	120
CAGATACTAA TTGAAGATTT AATACTTCTG CTGCCATAGC TAGCTCCTCC TATTTAAATT	180
TTTGGGATTA AGTACTTTAT CCACCCTCAT ATACTCTCTC CACCAGTAAA ATGCAAGCAA	240
TGATACAAAA TAGATTTAAC TATTTTATAT AGCGAAA ACT TACAAATTTT TAAGAAATAA	300
TTTTTGCA TT CTTAAAGATA AAATAGGAAC TTTTAGTAAT AAATATTAAA ATAAATAAAA	360
TAATAGATAC TATAAAATTT GGAAGTATTA ACCCCAAAAG ATTCATATCA TCTATTAAAA	420
TATCCTCTAA AGAGTAGTAT ATTAAAGCCA TAATTTTAAT GTTAAGTAAA AATGCAATTA	480
ATGAAGTAAC AAATGTCAAA AATATAGCCT CACCAACTTT AATCTTAACC ATCTGGTAAT	540
TAGAAGTTCC TAAAATTTCA AATTGCTGAA TCTCAATCCT TTCTTGATGC GATGACAAAA	600
ATGCAATTGA AATAATATTT GCAAGTACTA TCAAAATTGG TGCTCCTACA TAGACAATAA	660
ATGCTACTTT TAGCTCTAAA TCACTGTCAT CTTGAAATTG AGATAGTATA TTCTGAGAAA	720
TCATTGAAA ACTAGAAATT AGTAATATAG CTCCTGTAAT TGCAGCACTG ATAGATTTTA	780
TATAAGACTT ACAATATAGT AAATTCCACT TCGAAACAAT GAACATAAAA TTATTCTTAA	840
ATATAATTAT AGAAAGTAGT TTGATAAAAC ATGACTGTAT AAAAGGAGAT AATTGATAAA	900
TAATCACAAT ATCTAAGATT ACAATATTGA ATATTATCTG GGCCTTCGCT AAAATGTGTC	960
TATCTTGAA AATTGTGTC AAAGAAAGCA ACCAGATAAC ACTAAAACCA GCCAATAGCA	1020
GTATTCTTTT TACTATTGAA AGAACATGCC TTATTTTAGA ACTCTTCCTA TTTCTAATCT	1080
TCTTGAACGT ATAAAAGCAA CCACTTAGAA AGGCTAAAA TGAAATCAAC ACTACTGTAA	1140
TGATACATCC AACAGCACTC GTTTGAAATT GGATATCAGG TAATATATTT TCCCCGAAAA	1200
AGTATTGTAA AAAATAATAA TAATTGACG TAACAAATAT AGAGCATAGA TATGCAATAA	1260
AACTAATAAT CGAGGAAATG ATAAAAATCT GTCCCCCAC AAGAAATGAT AGTTGAAGGC	1320
GACTTGCTCC CAACACCTCC AGAAGTTCGT AATCATCTCT AAAAATTTCA ACCAACATAT	1380
TTATTATGTT AGAGAGCACA AAGAATAATG TTA CTCTCC GAATACTATC GGAAACATAA	1440
AAATPGGTTT AGGATCTGGA AGTCCGACAA ATACTTGCGA ATTATTCTCA ACATTAATTA	1500
CCCCATTAAC AGCCAATCCC ATA ACTAAAC TCGAAACAAA AATTACTGGT GAAACGCCTA	1560

407

ACCATTGTTT CTTATTATGT AAAAATTGAT AGTAAACTAA TCTGAGCATC TCTATTCCTC	1620
CGTAGTTGAT TGTACCTCTA AGATTTTATA CAACTCTTCC CCGCTAGGTC TATGAAGTTC	1680
TTTGAAAATT TTTCCATCTT TCAATATTAA TGCACGATCA GTTTTCGAGG CCAATTCTAT	1740
ATCGTGCGTT ACCATAATTA CACACTTACC CGCCCCTACT AACTCTCTCA ATAATTCAAA	1800
AATTACTTCA CGAGAAACGC TGTCTAAAGC CCCAGTTGGC TCATCAGCAA ATATTATATC	1860
ACTATCAGCA ATAACCGCTC TAGCTATAGC AACCTTCTGT TGTCTCCAC CAGACAGAGT	1920
TCCAACAAAA TCGTTTAAGC CAGCATTAAT CTTCATTCTT TTGAGTAAAGT TTTCTACATT	1980
TTTAATAGTT AATTTTTTTT GTGATAATCG CAAAGGAAGT GCTATATTTT CTATTACCGG	2040
CAGGGAAGGT ATTAAATTGT ATGCTTGAAA TATAAAAGAT ACTTCGTTAC GTCTTATACT	2100
TGACAATTTT GCATTTCTGA TTTTATAGGG GTTGATTCCA TTAAATTA CTTCCTCACT	2160
TGTTGGTTCA AGCAAACTAG AAATACATTT TAATAAAGTT GACTTTCCAG AACCACTAAT	2220
TCCTAGAATA CTTATAAATT CTCCTCTCGA AGCAGAAAGA GAAACATTTT TCAGCACTTG	2280
CAACGTTTTA TTATTTCCTA GTAAAAATTG ATGATACAGC CCTTTCACCT TTAATATATA	2340
ATCTTTATCC ATATTCTTGC CTCCAATCAC TTAATTTTGA AAAGTGTTC ATTTTCCAAT	2400
TTATATATAT CAGTGATCT CTGTGATTT AAGTCATAAT GATGTGAAAC TTCAATAAAT	2460
GAAATACCTA AATTGAACAG AATATCATGT ATGGAATTG AATTATCATT ATCTAAATTA	2520
GCTGATATTT CGTCAAATAA GTACACTTTA TTATTCTTAA TCAGAGCTCT AGCTAAAGCT	2580
ATTTTTTGT TTTGACCTCC AGACAAATTA CTACCATTTT CACCACATTG ATAATTTAGT	2640
ATATCTATCT TTTCTAATTC TTCATATAGA TTTACCTTTT TTAACACCTC AATTATCTGA	2700
TCATCTGAAA AATATTCATT TTGAAATAAA GTTACGTTCT CACGAATAGT AGTGTCAAAA	2760
ATATATGGTG TCTGATCAAC TGTGGTATT GAATCTGAAC TCTTTTCCC ATGTGATAAC	2820
AAATTACAT AACCTTTTTG TGGCTTTAAA GAACCATTAA TTAAATTTAA AATCGTTGTT	2880
TTCCCACTAC CAGAAGTTC TGTAAATAAT ACCCTAAATG GTGACTTAAA TGAGAAGTCA	2940
ATACTTAATT TATTTTCTGG TGAATAGAA TATACAACAT CTTTCATGTG TATCTCATCT	3000
ATTGATGAAG TATACAGTCC GTTATTATCA TGTTCAGCGT CTATAAAAT CTCTCTCCA	3060
CTTAAGTATT TTA AAAACGG TTTCCTTAAA TCTTTGGTTG TATTTATCTT ATTTAATGAA	3120
TAGGCAATTG ATTGTATCGG CCCTAAACT TTATCGTTTG CTAAGAAAAT ACCTATCAGT	3180
TCCTAAAAG AAAGGCTTTT ATGATAAATT ACAAATAAC ATCCTACAAC CAAGGGAAGT	3240
AGAAAGCAA AACCTGAAAT TAGTACTGCA ACCAATTTTG AAAGAACCTC TGATCGTTTC	3300

408

AAATTAAAAG TAGAATCTTC TAGTTTATCC AACTTTTAT CCGACAAACT AATTATTTCT	3360
TTAGTAACAG AATAAGATTT TAATGTCTTA AAACCATTA AAATTTCTTT TATTATGTGA	3420
GTATACTCTG CATTGCTGTT AGAGTACTCA TTAGCTGAAT TAGACAACAT CTTCTTCATA	3480
AAGACAGGTA CTATAATCGG CAATGCTGAT AATACAATAA ATATTATTGA NACTAGGAAG	3540
TTTAAATAAA GCATAAACT TAGAGAGACG ATGAACAACA ATATTGAAGA AATTATTTCA	3600
AAAATTTGTC TAAAATAGTT TTCTTCGATT AATCTCAAAT CATTTGACAA AACTGAAATA	3660
ATAGATGAGT AATCTTTAAC CATTTAGAA GAAAGATACT GTTCTCTAAA ATATCCTTGT	3720
TTAATTTTTA CATTTATATC TTTAGTTATT GATGCTCCG TTACTTCTAA ATAGTAATTT	3780
GATATATAGA TTGCTGACCA ACCCAGAATA CTTATAGCAC CAAATCTTAG AACGTCAGAA	3840
AATGAGGAAG TCTGATTTAA ACTACCTGCA TATACAATAA TTCCTGAGAG CAAGACACCA	3900
TTAAACGAAG ATAGAAATAT TAAAATCCCC ATTAATATAA GTTTAGTCTT TTTTATAAAT	3960
TTTAAATAAT TCATAAGTTA TTCCTTCCCA CTTCTTCAA GAAATAATTT AAAGTATCAA	4020
TCATTAAGAG AACATCTGAT GGAGTAAAC CTCCATGACC AGCTGCTTTG TTTAAATACA	4080
ACAAACTTTT AACTCCAATA GAATTTAATT TCTTTGACCA CTCTATCACT TCGTTATTAT	4140
TAATATATGG GTCTTTCTCA CCAAAATAT TAACTATAAC AGTATTTGAG TCTCGTGCCT	4200
TTTCAATATT TTGCATAGGC GAATATGACT TTATATAAGC CTTTACTTCA GGGTCTCTAA	4260
TATCTCCCA CTCTGCTATT TCGGTCTTAG AAAGAGGATC ATTTGGATTC TGAAGTGAT	4320
CATAAGGATT TATAAATGGC GAAAATAAGA GAATGCTTTG CAATAAATTT TTTTCTCGT	4380
TCAACACCGC ACCAGCAATT ATTCCACCTG CACTAGAAGT TATTAAACCT AATCGCTTAC	4440
TGTCAAATAC ATCATTTTCC CTTAAATAAT TTAATCCCTC AATAAAATCT CTGATAGAAT	4500
TCCATTTGTT TAACGCCTTT CTGAGCGAT ACCATTCAAC ACCCAAATAG CCTCCACCTC	4560
TTACATGAAC TATAGCATAA ATAAAACCTG CATCTATTAT AGATAACATA ATTCATCTA	4620
AATCAGAATT ATCATTTCTTA CCATAAGCCC CATAGACACT TAGAATACAT TTTTCTCTC	4680
TTGGGAGCTC ATCCGTATCT TCACTTTTCC AAAATAAAGA AATCGGTATG CTTACATCAT	4740
AACTGTCTTT TTTAGTCCAA ATCACCTTAG AAAAATATTT AGTATTATTC GATTTTATGA	4800
TGGGTCTTTC AAATTCAGTT TTTAATGTAT TTTCTATTAA ATCAAACTA AGTATTTT	4860
CGTAAAAAGT TCTCCTCTCT AAAAACAGAA GAACACGATC AGAAAAAGAA TTTTCATAAA	4920
GTGTTGTCTT TTCATCAAAT GTTATCTTAT TAACACTCAA CTCCCTCAA CTATTATTTT	4980
TAAATGTAGC AAGATAAAG ACGGAATTCG CTGCGTTTGA ACAGTCTAAA AGGATATAAC	5040
GTCTATACA GTGAACCTCT CTAGCCCTAT CTTGATATGG TATAGTAATA GAACTCTGT	5100

409

CTCCCGAAGA AGTTTCCCTT AGAATTAGTT GATCTTTCTT TTCTTCAGTT GAAGAGAGCC	5160
CAAGAAAGTA CTGTGCTTTT TCTGTAATAA ATAGAGCGAT ATCTCTAGGT GTTGGGGCTA	5220
CCGTTTCTGT GTAAGAGTGT CTAACAAAAC CCGTCCGGTC GAAACTGTAT AGAAAAATCC	5280
TGCCTTTCTG AAAGTCTACT GACTTTACAA AACAATTATT GCTATCAATG TGGACTATTT	5340
TTAATCGAAA AGAGCATTCG TTTTCTTCAA ACAGTTCCTC TTCTGTAAAG CTATCAAAAG	5400
ATTATAGAAA TAACTTACTT GGCCTCCCGT ACTCTTTGGA GCGAGTATAC ATAACACCGA	5460
ATTTACCCAA ATAGAACGAA CTTTCTACTG AAATATCTTC AATGATAAAT AACTCTTCCA	5520
TAGTATATTT TTTTATTCCA ATTAAATTAG TCGTACGCAG TGAGGATACA ACCAAAATA	5580
TATAACTCTC ATCAGATGAA ATCCTAACAT CCTGTAAGAT ACTATCATCT GGCAGATAT	5640
ATTTTCCAC ATCAAGACA ATTTAAGTG AATTGAATT GTCTAACTG GAAGAACTAA	5700
CCTTAGGAAT CCAGTCATTA TCTTCGACAT ACCATTCTT TATTACACCA GTATTGGTA	5760
TACTCCAATT ATCAAATTGG TACCAATATC GCCCTCTCCT AAATATCAAA GAATTCCATT	5820
TTTTTAATTC CTGAAATGAT GAAGAGATAG ACCTCTTATA GTGTGTTTTT TCCTGTATTG	5880
TATTTAAAAA TATTTTATTA CTCTGATTCA CAAGTATGAC CCCTTAATAA TGGTATCTAA	5940
ATATTATATT TGAGGAAGAA TCGTCAATTT ATTATCCATT ATTGATACCA ATCCAATTGC	6000
AACACCCGCA AATCCCGAAG CAATATCTGT TGTATCTTTT AAACCATTAT CTCCCGCAAT	6060
AACAAATCCT TCTTCAATTA CACACAAATA TCTATAAAGT TGTTCATTA ATTTCTTTTG	6120
TCCTGAAAAG TTATCATCGA TATCACTATA TATATTATTA GCAACTTCAA GACCACAAA	6180
TCCGTTAAAT AAACCTGGTA ATACACAAA AACTACATCA GTTGCCCTCT CTAAAGAAGT	6240
TAAATATTTT AAGTATTTGC TTGACAAGAT TTCTTTATTT CTATTAATAA GTAAAAGCAG	6300
GCCAGCACTT CCAGTTGCTA GATATGGTAG TAATCTATGA CCTTGGCTGT ACTGCAATGA	6360
ATTATTACTA TCTACTTTAT AAGCACTAA TTCTTTATCT ACAGCCAATT CTAGACCATT	6420
TTTATAGATA CTTTACCAG TTAATTTATA AGCTTCACCG AAGAGCCAAG CTACCCCTGC	6480
GTGACCATAT AGTAATCCAC CAAAATCTC ATAAGGATCG TTAATCTGAA CATCACTAGC	6540
GCCAACTTTA CAAAAGTTT CTGGATTTT TATATAATTT AAAGTATATT CTCTAAGCCT	6600
AATTAGTATT TCTTCTCTA GTTTATTATC AATTCCCCCT TTAATAAGAA AATACAGTCC	6660
AACCAGTAAA ATTCCAGCCT GCCCACTATA TAAATTTTAA TTTTGTGAAT TCTCAAATAT	6720
CTCTATAAAA TGAGTTGTAA AAAGTCAAC TGCCCGATCT ATCTCCCCAA ATTCATAAAT	6780
GAGCCAGATT GTACCAATTT TACCATCAA AAGACCAGAA AGGGACGATT TCTTAAATTT	6840

410

ATTTACTGCC TCATTAATAA CCTGTGTTTCG AATCTCATAA TAGTCATCAA ACTTGAAATT	6900
TTTTACTTTC TTAGCTAGTT GTTGATAACT CCAAAGGATA GCTAAATCTG AAAACGCAAT	6960
TCCTTGATTA AAATTCAGAC CATAATAATG AACTGGGAAG AATCTTGATT GAAATTCCTT	7020
ACGCCACTGT CCATAAGTTA GCGTAAACCC TCTCAATAAT TTTATAATAA AATCTTGAT	7080
ATCTTGCTCA CTCTCGATAG TTCTAATCTC ATGCATGGGT TTTAAACTT TTTCTCTGGA	7140
AATATTCTCA ATCTGTGGAC ATTTAGAATC TAGATATGAC AATAAACTTT CTACATAATC	7200
TATATGTTCT CTGTATAAC CCAAAGACTC AAATAGTTTT TTTCTCTCTA TCCTGGTTTG	7260
ACTTACATAG TTGTATGTCA AATCCGATGT AGTTACTAGT GGCATGTATA AATAATGAGC	7320
TATTTGTCTA ATACCATACC AATCTATCTC ACTGGGAAGT GTTCTCTGCC ATGCTCTAAA	7380
ACCAGGGGCT GCAACTTTAT GTACAACCTT TTCATCATTT GAAAAGACAG CCTGTTCCCA	7440
GTCTATTATA CTAATCTCAT CTTATCCTT AACCAAGATA TTTCTTAAAT GTAAATCTTG	7500
ATGATATACA TTTTCAGAAT GAAACTTATT CGTTAAATCG ATGAGTTTTT CTACTATCTT	7560
TGAAACTCTC AATAGATAAT CTTTGGTCTT ATCAACAAC TATATAAAG GAAATATT	7620
GGTAACCCAT CTATTTAGTG GAACGCCCTT CATATGTTCA ATTCCTAAGA AGGTGTGCTC	7680
CCAGATCTTA CCGTGCCAGT ATATTTTAGG CGTCTCACTC CATTCAATTA GAATTTTAG	7740
TGCTTTGCAC TCCGAAGCTA ATTTCTCTGA AGAATAAGTA CCATCAAATC CTAGACCTGT	7800
ATACGGTCTA GCCTCTTTTA AAATTATTTT TTTCCCATCT TCTTTTAGCC TAGCATATA	7860
TATCCCACCA CTGTTTGAAA ATCTAATGTC ATTATCTATA ATAAAGGGAA AGTCTCCCTG	7920
TTTTTTATCT TTCTTGTCAA GCCATTTATT CAAAAAGTCA GGGGGCACTA TACCTTTTGG	7980
AATTTTAAAT ACTGGTAAAC GTTCATCTTT AACAACTTCA TCGCCAACAA TTAATTCATC	8040
AATAGCAACC TTCTTTTCAT CATCCCTTGA CGGCCTAAAC ACACCATACC TCAGATATAT	8100
TGGTGCTTCA TCCCAACGTT TATCGCTTAA AATATATGGC CCATTATATT GCTTTAAGGC	8160
ACTTTCTAAC CTTTGCAAAA CCGACTCTAA TTCATTTTGA TTTGGATAAC ATGTAATAAA	8220
TTTACCAGAA AATCCTCGAC TAACCAATTT CCCGTTTCGC ATGATAAATT TGTCTTCTGT	8280
ACTAAGATGT TTAATGGAA TTCGCATTTT ATGGCAAATT TTTGCTACAT CTTGTAACAA	8340
TTCATGTGAA CTGTTATACT CTGAACCTAAT GTGTATTTT CACCCTTGTC TTTCAACAAA	8400
TTTTCCAATA GGGTATTGAT AAACCCACTC ATCATTATTC ATTACTTCGT GCCAATTAAA	8460
AGGCAGACTT ACTTGGTACT TTATGCTAGT ATCTGTACTA TAATCATTAT TAGTGAAAAA	8520
GAAAGGATGC TCCAAATTGA AATTATAATC CATAACAAA TCTCCAAGAA ATTTTATCAA	8580
ACTTAATATA TCTATAGCTA GACAGACTTA TTAAATAAA AAGGGAGAAT CCTTTGGATT	8640

411

CTCCCCATAT AAGCACTAAC ATTCCAACGT GCACATATTG GAACGACATC CATAACTCCA 8700
 GAGAATCTCT AAAGTTTACA ATTTAAATGA ATTAACAATT TTCCCAACTA AAAGCACTCC 8760
 AGTTACCGCA ACGATTGTGTA CTGAATGTAC TAAATCGCAT TCCATCAACT TCATCTGTTT 8820
 CGTCAACTTG AACAGATACT AATTGAAGAT TTAATACTTC TTCTGCCATA GCTAGCTCCT 8880
 CCTATTTAAA TTTTGGGAT TAAGTACTTT ATCCACCCTC ATTATACTCT CTCCACCAGT 8940
 AAAATGCAAG CAATTATACA ATGTTGTCAC ATAGAAAATA ATGTTTCCGT AACTTTTCAA 9000
 AGTAACTTCC ATCTCTCTCC CAAAACCTGA AGTTAGTTTT AGAAGTTACC TAAAAATCAG 9060
 GTCACCTATT TAAAAAAGC AGCAAACCTAT AACTAGTAG GTTCCACACC AAATGTAGTC 9120
 CCATACTGCC CCATAAGTCA GATTTATAGC GCACCATACC TAAAAACATC CCAAGTGAAA 9180
 CATACAAACA CCAAGCTAGA ATGGTTCCTG TATGATGTGC TAAGGCAAAT AAAACACTTG 9240
 TCAAAGCAAC TCTGATATCT AATTTTCTGA CCAAATTTCA TAAATTTCT CGATACAGAA 9300
 ATTCTTCAAC CATACTCGCA TTGATTAAGA ACAATAAAAA TGAAAACCAA GGAATTTGAT 9360
 GTTGAAGGCC AATTAAGTTT GCTTGATTCTG TGCTTCCTTG AGCATGAATC AGACTAAAAC 9420
 ATAGACTTAT AATCAGTAGG CTAACAAAT CAACACCAAG CCATTTCTATC CTAGATTTC 9480
 TATTGACCTT ATGCGCTTGT TTGCGTTGGC CATACTCCA TAAAAAGAA ATGAGTGACG 9540
 AACCATAGAG AATCTGTAGT ATAGTTACT CACCGATACA AAGAAATTTT AATAAGTATA 9600
 GAGTACCAA TAsGACATTT ACTTGTGGA ATATATAAAC TGAATTATT CTTTTCATAG 9660
 TTACCTCCGA AATAAATCTT CATAATCTAA ATCTAATACC TGCACAATCC TTT 9713

(2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8657 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

AAAGAAATTG TCAGAGAGTG GCTAGATGAA GTAGCAGAGC GGGCTAAGGA CTATCCAGAG 60
 TGGGTGGATG TTTTCGAGCG TTGCTACACC GATACCTTGG ACAATACGGT TGAAATCTTA 120
 GAAGATGOTT CAACTTTTGT CTTGACTGGG GATATTCCTG CCATGTGGCT TCGAGATTCTG 180
 ACAGCCCAAC TCAGACCCTA CCTTCATGTA GCTAAAAGAG ATGCCCTCCT GCGTCAGACC 240
 ATTGCAGGTT TGGTCAAACG TCAGATGACC TTGGTACTCA AGGATCCCTA TGCTAACTCC 300

412

TTCAACATTG AGGAGAACTG GAAAGGGCAC CACGAGACTG ACCACACAGA CCTTAACGGC	360
TGGATCTGGG AGCGCAAGTA TGAGGTGGAT TCGCTTTGCT ATCCTTTGCA GTTGGCTTAT	420
CTCCTCTGGA AAGAGACTGG CGAGACTAGT CAGTTTGATG AGATTTTGTG CGCAGCGACT	480
AAGGAAATTC TCCATCTGTG GACGGTGGAA CAAGACCACA AGAACTCTCC TTATCGTTTT	540
GTCCGAGATA CGGACCGTAA GGAAGACACC TTGGTAAATG ATGGCTTTGG ACCTGACTTT	600
GCAGTGACAG GTATGACTTG GTCAGCTTTT CGTCCGAGTG ATGACTGTTG CCAGTATAGT	660
TACTTGATTG CGTCAAATAT GTTTGCTGTA GTAGTCTTGG GTTATGTGCA AGAAATCTTC	720
GCAGCATTA ACCTAGCTGA TAGCCAGAGT GTTATTGCTG ATGCCAAGCG TCTTCAGGAT	780
GAAATCCAAG AAGGAATCAA AACTACGCT TACACCACCA ACAGCAAGGG CGAAAAGATT	840
TACGCTTTTG AAGTGGATGG CCTAGGAAAT GCCAGCATCA TGGATGATCC AAATGTACCA	900
AGTCTACTAG CTGCGCCCTA TCTGGGCTAC TGTTCGGTCG ATGATGAAGT GTATCAAGCT	960
ACTCGTCGTA CCATTTTGAG CTCTGAAAAT CCATACTTCT ACCAAGGAGA ATACGCAAGC	1020
GGTCTCGGCA GTTCTCATAC CTTCTATCGC TATATCTGGC CAATCGCCCT TTCTATCCAA	1080
GGCTTGACAA CAAGAGATAA GGCAGAGAAA AAATCTTTCG TGGATCAGCT GGTTCCTGTC	1140
GATGGTGGTA CAGGTGTCAT GCACGAAAGC TTTCATGTAG ATGATCCGAC CCTCTACTCT	1200
CGTGAATGGT TCTCCTGGGC TAACATGATG TTCTGTGAGT TGGTCTTGA TTAATTGGAT	1260
ATTCGCTAAG GGGCTCGCTT TAGCTCAACC GATTCTTATC AGAATCACAA GTTTACATTT	1320
AAAACGTAA AATTAAAT TAGAATGAGG TTTTACTTCA TGGAAAATGT TGTGTACAT	1380
ATTATCTCAC ATAGTCACTG GGATCGTGAG TGGTACTTGC CTTTGTGAAAG CCATCGTATG	1440
CAGTTGGTGG AATTGTTTGA CAATCTCTTT GATCTCTTTG AAAATGACCC TGAGTTCAAG	1500
AGTTTCCACT TGGATGGACA AACTATTGTC CTTGATGACT ACTTACAAAT TCGCCCTGAA	1560
AATCGCGACA AGGTCCAACG CTACATTGAC GAGGGCAAAC TTAAAATTGG TCCCTTTTAC	1620
ATCTTGCAAG ATGACTACTT GATCTCCAGT GAAGCCAATG TCCGCAATAC CTTGATTGGT	1680
CAACAAGAAG CTGCCAAATG GGGTAAATCA ACCCAGATTG GCTACTTTCC AGATACCTTT	1740
GGAAATATGG GACAAGCGCC TCAAATCTTT CAAAAATCAG GCATTACCGT GGCGGCCTTT	1800
GGTCGTGGTG TGAAGCCGAT TGGATTTGAC AACCAAGTCC TTGAAGATGA GCAGTTTACG	1860
TCTCAGTTTT CAGAAATGTA CTGGCAGGGT GTGGATGGTA GTCGTGTTTT AGGTATTCTC	1920
TTTGCCAACT GGTACAGTAA CGGGAATGAA ATTCCAGTTG ACAAAGATGA GGCCTTGACC	1980
TTCTGGAAAC AAAAATTGTC AGATGTGCGT GCCTACGCTT CGACCAACCA ATGGTTGATG	2040
ATGAACGGCT GTGACCACCA GCCTGTACAG AAAAATCTGA GCGAAGCCAT TCGTGTGGCA	2100

413

AATGAACTCT TCCCGGATGT AATCTTTGTT CATAGTTCTT TTGATGAATA TGTTCAGCT	2160
GTAGAAGGTG CGCTTCCTGA AACTTATCA ACTGTTACAG GCGAGTTGAC CAGTCAGGAA	2220
ACAGATGGCT GGTACACACT TGCCAACACT TCTTCATCCC GCATTTACCT AAAACAAGCC	2280
TTCCAAGAAA ATAGCAACCT CCTAGAGCAA GTGGTAGAAC CCTTGACTAT TATCACTGGT	2340
GGACACAACC ACAAGGACCA GTTGACCTAT GCTTGGAAAA CACTTTTGCA GAATGCGCCA	2400
CATGATAGTA TCTGTGGCTG TAGCGTGGAC GAAGTTCACC GCGAGATGGA AACGCGTTT	2460
GCCAAGGTCA ACCAAGTAGG AACTTTGTT AAAAGTAACT TGCTCAACGA GTGGAAGGGT	2520
AAAATTGCTA CGGATAAGGC TCAAAGTGAC TATCTCTTTA CTGTCAITAA CACAGGCTTG	2580
CATGATAAGG TCGATACTGT CAGCACAGTG ATTGATGTGG CGACTTGTGA TTTCAAGGAA	2640
TTGCACCCAA CAGAAGGCTA CAAAAGATG GCTGCTCTTA TCTTGCCAAG TTACCGTGTG	2700
GAGGACTTGG ATGGTCGTCC TGTAGAGGCT ACAATCGAAG ACCTCGGAGC TAATTTTGAG	2760
TATAATTTAC CAAAAGACAA GTTCCGCCAA GCTCGTATFG CTCGTCAAGT GCGCGTGACC	2820
ATTCCAGTTC ACCTAGCGCC GCTTTCTTGG ACAACCTTCC AATTGCTGGA AGGAAAACAA	2880
GAACACCGTG AGGGTATTTA CAAAACGGA GTGATTGATA CACCATTCTG AACGGTGAGT	2940
GTGGATGACA ACATCACAGT CTATGACAAG ACAACTCACG AAGCCTATGA AGACTTTATC	3000
CGCTTTGAAG ACCGTGGGGA CATCGGAAAC GAGTATATCT ATTTCCAACC AAAAGGAACA	3060
GAGCCAATCT TTGCAGAGCT TAAGGGCCAC GAGGTCTTGG AAAACACAGC TTGCTATGCT	3120
AAAATCTTGC TCAAACATGA ATTGACCGTG CCTGTCACTG CGGATGAAAA GCTAGAAGAA	3180
GAGCAACAAG GTATCATCGA GTTTATGAAG CGTGAGGCTG GACGGTCAGA AGAATTGACA	3240
AACATTCCTC TGGAAACTGA GTTGACTGTC TTCGTTGACA ATCCACAAAT CCGCTTCAAG	3300
ACTCGCTTTA CTAACACTGC CAAGGATCAC CGTATCCGTC TCTTGGTCAA GACTCATTAAC	3360
ACGCGTCCAA GCAATGATTC TGAAAGTATC TATGAGGTGG TGACACGACC AAACAAACCA	3420
GCTGCTTCAT GGGAAAACCC TGAAATCCT CAACACCAAC AAGCTTTTGT CAGTCTGTAT	3480
GACGATGAAA AAGGGGTGAC TGTATCCAAC AAGGGATTGA ATGAATACGA AATCCTTGGG	3540
GATAACACCA TTGCCGTGAC CATTTTTCGT GCATCAGGTG AGCTAGGTGA CTGGGGCTAC	3600
TTCCCAACGC CAGAAGCACA ATGCTTTCGG GAGTTTGAAG TCGAGTTTGC ACTTGAATGC	3660
CACCAAGCCC AAGAACGCTT CTCAGCCTAT CGTCGTGCCA AAGCCTTGCA GACACCGTTT	3720
ACCAGCCTTC AGCTTGCTAG ACAGGAAGGA AGCGTGTTG CGACTGGTAG CCTCTTGAGC	3780
CATTCTGTTT TCAGCATACC GCAAGTTTGT CCAACAGCCT TTAAGGTAGC TGAAAAAGAA	3840

414

GAAGGCTATG TGCTTCGTTA CTACAATATG TGTAGTGAAA ATGTACGTGT GCCAGAAAGT	3900
CAACATCTCT TCCTTGACCT ACTTGAACGA CCATACCCAG TTCATTTCAGG ACTATTGGCT	3960
CCACAAGAGA TTCGTACAGA ATTCATCAAA AAAGAAGAAA TTTAATTTCA AAAAGTAAAC	4020
ATCAAAAGAA AGGAGGGGCG AAAAAGTAAG AACTAACTGC TGATTTCGCCC CTTTATGGT	4080
AAAAACAATG ACCATTGCAA CGATTGATAT CGGAGGGACT GGGATTAAAGT TTGCCAGTCT	4140
GACTCCTGAT GGGAAAATAC TGGATAAGAC AAGTATTTCA ACGCCTGAAA ACTTGGAGGA	4200
TTTACTAGCG TGGCTAGATC AACGCTTGTC AGAACAGGAT TACAGTGGGA TTGCTATGAG	4260
CGTTCAGGT GCAGTCAATC AAGAGACAGG TGTGATTGAT GGCTTCAGTG CGGTGCCTTA	4320
CATCCATGGC TTTTCTTGGT ATGAGGCGCT TAGCTCTTAT CAGCTACCTG TCCATTTAGA	4380
AAATGATGCC AACTGCGTTG GACTCAGTGA ACTACTAGCT CATCCAGAGC TTGAAAATGC	4440
AGCCTGTGTC GTGATTGGGA CAGGGATTGG CGGAGCCATG ATTATCAATG GTAGACTTCA	4500
TCGAGGTCGC CACGGTCTGG GTGGAGAATT TGGCTACATG ACAACCCTTG CCCCTGCTGA	4560
AAAACTTAAT AACTGGTCGC AACTAGCATC AACTGGGAAT ATGGTACGAT ACGTGATTGA	4620
AAAATCTGGT CATACTGATT GGGACGGTCG CAAGATTTAC CAAGAGGCCG CAGCTGGTAA	4680
TATCCTTTGT CAAGAAGCCA TTGAGCGCAT GAACCGCAAT CTGGCGCAAG GCTTGCTCAA	4740
TATCCAGTAT CTGATCGATC CAGGTGTCAT CAGTCTGGGT GGCTCTATCA GTCAAAATCC	4800
AGATTTTATC CAAGGTGTCA AGAAGGCTGT TGAAGACTTT GTCGATGCCT ACGAAGAATA	4860
CACGGTCGCA CCAGTTATCC AGGCCTGCAC CTATCACGCA GATGCCAATC TCTACGGTGC	4920
TCTTGTC AAC TGGCTACAGG AGGAAAAGCA ATGGTAAGAT TTACAGGACT TAGTCTCAA	4980
CAAACGCAAG CTATTGAGGT TTAAAAGGT CACATTTCTC TACCAGATGT GGAAGTGGCT	5040
GTCACTCAGT CTGACCAAGC ATCTATCTCT ATCGAGGGTG AGGAAGGTCA CTATCAATTG	5100
ACCTACCGCA AACCTACCA ACTTTATCGT GCCTTGTCCT TGTTGGTAAC AGTTCTAGCA	5160
GAAGCTGATA AAGTAGAGAT TGAGGAACAA GCAGCTTACG AAGATTGGC TTACATGGTT	5220
GACTGTCTC GAAATGCGGT GCTGAATGTG GCTTCTGCCA AGCAGATGAT TGAGATATTG	5280
GCTCTCATGG GCTACTCAAC CTTTGAGCTT TACATGGAAG ACACCTACCA GATTGAAGGG	5340
CAGCCTTACT TTGGCTATTT CCGTGGAGCT TATTCAGCAG AGGAGTTGCA GGAAATCGAA	5400
GCCTATGCCC AACAGTTTGA CGTGACCTTT GTACCATGCA TCCAGACCTT GGCCCACTTG	5460
TCGGCCTTTG TCAAATGGGG TGTCAGGAA GTGCAGGAGC TCCGTGATGT AGAGGACATT	5520
CTTCTCATTG GCGAAGAAAA GGTTTATGAC TTGATTGATG GCATGTTTGC CACGTTGTCT	5580
AAACTGAAGA CTCGCAAGGT CAATATCGGG ATGGACGAAG CCCACTTGGT TGGTTTGGGA	5640

CGCTACCTGA TTCTGAACGG TGTGTGGAT CGTAGTCTCC TCATGTGCCA ACACTTGGAG 5700
CGCGTGCTGG ATATTGCTGA CAAATATGGT TTCCACTGCC AGATGTGGAG TGATATGTTC 5760
TTCAACTCA TGTGACGGA TGGCCAGTAC GACCGTGATG TGGAAATTCC AGAGGAACT 5820
CGTGTCTACC TAGACCGTCT CAAAGACCGT GTGACTCTGG TTTACTGGGA TTATTATCAG 5880
GATAGCGAGG AAAAATACAA CCGTAATTC CGCAATCATC ACAAGATTAG CCATGACCTT 5940
GCATTTGCAG GGGGAGCTTG GAAGTGGATT GGCTTTACAC CTCACAACCA TTTTAGCCGT 6000
CTAGTGGCTA TCGAGGCTAA TAAAGCCTGC CGTGCCAATC AGATTAAAGA AGTCATCGTA 6060
ACGGGTGGG GAGACAATGG TGGTGAACT GCCCAGTTCT CTATCCTACC AAGCTTGCAA 6120
ATCTGGGCAG AACTCAGCTA TCGCAATGAC CTAGATGGTT TGTCTGCGCA CTTCAAGACC 6180
AATACTGGTC TAACGGTTGA GGATTTTATG CAGATTGACC TTGCCAACCT CTTACCAGAC 6240
CTACCAGGCA ATCTCAGCGG TATCAATCCC AACCGCTATG TTTTATCA GGATATCTTT 6300
TGTCCGATTC TTGATCAACA CATGACACCT GAACAGGACA AACCGCACTT CGCTCAGGCT 6360
GCTGAGACGC TTGCTAACAT TAAAGAAAAA GCTGGAACT ATGCCTATCT CTTTGAACT 6420
CAGGCCCAGT TGAATGCTAT TTTAAGTAGC AAAGTAGATG TGGGACGACG CATTCGTCAG 6480
GCCTACCAAG CGGATGATAA AGAAAGTTTA CAACAAATCG CCAGACAAGA ATTACCAGAA 6540
CTTAGAAGCC AAATTGAAGA CTCCATGCC CTCTTTAGCC ACCAATGGCT GAAAGAAAAC 6600
AAGGTCTTTG GTTTGGATAC AGTTGACATC CGTATGGGCG GACTCTTGCA ACGCATCAAA 6660
CGAGCAGAAA GCCGTATCGA GGTTTATCTG GCTGGTCAGC TTGACCGCAT CGACGAGCTG 6720
GAAGTTGAAA TCCTACCATT TACTGACTTC TACGCAGACA AGGATTTTCG AGCAACTACA 6780
GCCAACCACT GGCATACCAT TGCGACAGCG TCGACGATTT ATACGACTTA ATATTCTTCG 6840
AAAATCTCTT CAAACCACGT CAGCTTCCAT CTGCAACCTC AAAACAGTGT TTTGAGCAAC 6900
CTGCAGCTAG CTTCTAGTT TGCTCTTTGA TTTTCATTGA GTATAAAAAC AAGAACACCT 6960
TGCTTGGCGC AGGGTGTTTC GCGTGAAACA GAAGAATTAT CTGGTTTCAA ATGCTACAGT 7020
TAGACAACT TATGATAAAA TAGCAGAAAG TGAATGTTTC CTAAGAGCAA TTGGAGGTAT 7080
TATGCTACAC TTAATAATTAG TAAAACAAGA AATAGAAGCT GAAAAGCCAG CATCTGTAGA 7140
AGCTTGGATC ATTTCCGTCA AATTTAAAAA AGGTTGCTAC CGACATATAT AGATTCCAAA 7200
AACAAAAACG TTAGCGGAAC TAGCAGATGT GATTTTATGG AGTTTGTATT TTGCAAATGA 7260
TCATGCTCAC GCATTTTCA TGGATAATGT TGAGTGGAGT CATGCAGATT CTTACTTTCCG 7320
TAGCTTTGTT AGTGACGATG TTGAAGAACG TTACACAGAA AATGTCTATC TGGATAGCCT 7380

416

AAGTGTCAAA CAAAAATTTA AGTTTATTTT CGACTTCGGT GATGAATGGC GTTTTGAATG	7440
CCAAGTGCTG AGAGAAATCG AGACAGAGGA CGAAGAAGCT TATCTCGTAC GTTCGGTTGG	7500
AACGTCGCCA GAACAATATC CAGATTATGA TGGTTTTGAC TATGAAGAAT GGTAATAATTG	7560
AAATCAGTCT GTGTAGGCTT AGTATTTCAA TAGACTTCCT GCAAACTAG AATCCTAGTT	7620
CATGATTGAT AATACCAGCA ATCAAATTCA TTCGTAATCC GAAGCGTTTA CGATGATTTT	7680
GATAGGTTGT TGAAAACATT TTAAACGTTT TTTACTTTGGC AAAGATGTTT TCAACCTTGC	7740
TTCTCTCCTT AGATAGCGCA TGGTTATAGG CTTTATCTTC AGCTGTTAGT GGCTTGAGTT	7800
TGCTGGATTT ACGTGAAGTT TGTGCTTGAG GACATATCTT CATGAGCCCT TGATAACCAC	7860
TGTCAGCCAA GATTTTACCA GCTTGTCGCA TATTTCTGCA ACTCATTTTG AACAACTTCA	7920
TATCATGACA ATAGTTCACA GTGATATCCA AAGAAACAAT TCTCCCTTGA CTGTGACAA	7980
TCGCTTGAGC CTTCATAGCG TGAAATTTCT TTTTACCAGA ATCATTGCTT AATTCCTTTT	8040
TTAGGGCGAT TGATTTTAC TTCCGTCGCA TCAATCATT CCGTGTCTC AGAACTAAGA	8100
GGAGTTCTTG AAATCGTAAC ACCACTTGA ACAAGAGTTA CTTCAACCCA TTGGCTCCGA	8160
CGGATTAAGT TGCTTTCGTG AATACCAAAA TCAGCCGCAA TTTCTTCATA AGTGCGGTAT	8220
TCTAGGCTTA ATTTAGGTTT TCGTCCACCT TTTGCGTGT TAAGTTGATA AGCTGTTTTT	8280
AATACAGCTA ACATCTCTTT AAAAGTCGTG CGCTGAACAC CAACAAGACG CTTAAATCGT	8340
GTATCAGTTA ATGTTTACT TGCTTCATAA TTTCGCAGGG AGTCTATTGA CTCTTTGGTA	8400
GGTGTCAATG TTTTTCAT CTATCCGAG AATTATTTT CCGCCATTG TATTTGCAAA	8460
TGCTGAGTAG GTTCCCAGA AAGACTCTGG AAGATTGTTT TTAGCTTTTT TGTATTCTAA	8520
ATCAACCCCT TCAAATTTTA AGTCCATATT TTTCCTTTAC ATCTGTTTTT TGTGGTTCTG	8580
GTATTTGTTT AAGTTGAGTG ATAATATAGC GAATTGAATT TCGAGAGTTT TTAATCAGTT	8640
AATTTCTTTT TTAACCC	8657

(2) INFORMATION FOR SEQ ID NO: 45:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11384 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

TCTATTTTGG GTATAGACTT ACCTATAAAG AAAATATCT ATACACTGCC TTACTAGCTA	60
TACTGAACGA GTCAACAAAA ACGATATATA TTGATGATAT AAATACAGCA AGATTTTTTA	120

ACTTCTTTGG CAATGATATT CCTAATTCGT CTTTAAAAA AATTGACTAT ATCGCACCTT 180
CAGAAATTGT TTCATTTAGT ACGTACGTTC GACAACGTTC TAAAGTAATT CCTAAAATTT 240
TGGAACATAT ATTAAATCA AGTTTTTTAT TAGAGAATAT AGATGTTTCT GGTTACACTG 300
TAAATATTTT AGAAGATCAA TTAACAAAAC ATAGAACAAT CAAAATTAGT AAAAATAAC 360
TGGTTGATCT CATGTATAAA TACCTAACAA AACCACGCGC CTTGCCTGCT GATGGAAGA 420
AAGGTACAAA TACATGAATA TCAAAGAAA AATCAAAAAG AATGGCCAAA GAGTTTATTA 480
TGCTAGTGTT TATCTAGGCG TTGACCAACT AACGGGCAA AAAGCCCGTA CAACTGTTAC 540
AGCAACCACT AAAAAGGGCG TTAAAGTAAA AGCGCGTGAT GCGATCAATA CTTTGTCTGC 600
TAATGGCTAT ACAGTTAAAG ACAAGCCGAC AATTACAACA TATAATGAGC TTGTAAAAGT 660
TTGGTGGGAT AGTTACAAGA ATACAGTTAA GCCAAATACT CGCCAATCCA TGGAGGGATT 720
GGTTAGAGTG CATTTATTGC CTGTATTGG CGATTACAAG CTATCTAAAC TTACTACGCC 780
TATTCTTCAA CAGCAAGTAA ACAAATGGGC TGACAAGGCA AATAAAGGCG AAAAAGGGGC 840
ATTTGCTAAC TACTCTTTC TCCATAACAT GAATAAGCGT ATTTTGAAAT ATGGCGTAGC 900
TATCCAGGTA ATACAATACA ACCCAGCTAA TGATGTCATC GTTCCACGCA AACAGCAAAA 960
AGAAAAGGCT GCTGTCAAAT ACTTAGACAA CAAAGAATTA AAACAGTTTC TTGATTATTT 1020
AGATGCTCTG GATCAATCAA ATTATGAGAA CTTATTGAT GTTGTCTGT ATAAGACTTT 1080
ATTGGCCACT GGTGCGGTA TTAGTGAGGC TCTGGCTCTT GAATGGTCTG ATATTGACCT 1140
AGAAAGCGGT GTTATCAGCA TCAATAAGAC ACTAAACCGC TATCAGGAAA TAACTCACC 1200
TAAATCAAGC GCTGGTTATC GTGATATACC AATAGACAAA GCCACATTAC TTTTACTGAA 1260
ACAATACAAA AACCGTCAAC AAATTCAGTC TTGGAAATTA GGCCGATCTG AAACAGTTGT 1320
ATTCTCTGTA TTTACGGAGA AATATGCTTA TGCTTGTAAC TTACGCAAAC GCCTAAATAA 1380
GCATTTTGAT GCTGCTGGAG TAACTAACGT ATCATTTTAT GGTTCGCC ATACACATAC 1440
TACTATGATG CTCTATGCTC AGGTTAGCCC GAAAGATGTT CAGTATAGAT TAGGCCACTC 1500
TAATTTAATG ATCACTGAAA ATACTACTG GCATACTAAC CAAGAGAATG CAAAAAAGC 1560
CGTCTCAAT TATGAAACAG CTATCAACAA TTTATAAAAA ATAAGGGTGA CCCATTTCCG 1620
GGCTACCCTC TTACTATACC AAAAATTAGT AGGGGTAGTA AAAAGGGTAT TAAATTATAA 1680
AAAGCACTAA GGGAAAGCGC CCCAAAGTGC TTATTTCAA GGCTTTATAG CCTATAATCA 1740
CATAAAGAGA TTATTTTTTA AGGTTGTAGA ATGATTTC AA TCCACGATAT TCAGCTACTT 1800
CACCAAGTTG GTCTTCGATA CGAAGCAATT GGTGTATTT AGCGATGCGG TCTGTACGTG 1860

418

AAAGTGAACC AGTCTTGATT TGTCTGCGT TAGTTGCAAC TGCAATATCA GCGATTGTTG	1920
AATCTTCAGT TTCACCTGAA CGGTGTGATA CAACAGCAGT GTAACCAGCT TCTTTAGCCA	1980
TTTCGATAGC TTCAAAAGTT TCAGTAAGAG TACCGATTG GTTAACTTTG ATAAGGATTG	2040
AGTTAGCAGC ACCTTCTTGG ATACCACGTG CAAGGTAGTC AGTGTTTGTT ACGAAGAAGT	2100
CGTCACCAAC AAGTTGTACT TTCTTACCA GACGTTTCAAGT AAGAGCTTTC CAACCATCCC	2160
AGTCGTTTTC ATCCATACCA TCTTCAATAG TGATGATTGG GTATTTGTTA ACCAATTCTT	2220
CAAGGTAGTC GATTGTGTTCT GCAGATGTAC GAACAGCAGC ACCTTCACCT TCAAATTTAG	2280
TGTAGTCGTA AACTTTACGT TCTTTATCGT AGAATTCTGA TGAAGCACAG TCAAATCCGA	2340
TAAATACGTC TTTACCTGGT ACATATCCAG CAGCTTCAAT CGCAGCAAGG ATAGTTTCAA	2400
CACCATCTTC AGTTCCTTCG AAACGAGGAG CGAATCCACC TTCGTCACCT ACGGCAGTTT	2460
CCAAACCACG TGATTTAAGG ATTTTCTTAA GAGCGTGGA GATTTCAGCA CCGTAACGAA	2520
GGGCTTCTTT AAATGTTGGC GCACCAACTG GCAAGATCAT GAACTCTTGG AAAGCGATTG	2580
GAGCGTCAGA GTGAGAACCA CCGTTGATGA TGTTTCATCAT TGGAGTTGGA AGAACTTTAG	2640
TGTTGAATCC ACCAAGATAG CTGTAAAGTG GGATTTCAGG GTAGTCAGCA GCAGCAGGAG	2700
CTACAGCGAT AGACACACCG AGGATTGCAT TCGCACCCAA TTTACCTTTG TTAGGAGTAC	2760
CGTCAAGTGC GATCATAGCA CGGTCAATAG CTGTGTTGATC ACGTACATCG TAGCCAATGA	2820
TAGCTTCAGC AATGATGTTG TTTACGTTGT CAACAGCTTT TGTGTACCA AGACCACCGT	2880
AACGAGATTT GTCACCGTCG CGAAGTTCAA CTGCTTCGTG TTCACCAGTA GAAGCTCTG	2940
ATGGAACCAT ACCACGTCG AAAGCACCTG ATTCAAGTGA AACTTCTACT TCAAGTGTG	3000
GGTTACCGCG TGAGTCTAGG ACTTCGCGAG CGTAAACATC AGTAATAATT GACATTTTTT	3060
ACTCTCCTTA TGAGTTAAAT TTTTACACC TCTATAATAC CTTAAAACCC CTCCTTTTTT	3120
AAGAAAAAAC GTTATCTTTG TGCAACTTTT CCTTAACTTT ATAAAGTAAT CGCTTTCTTT	3180
TGTCGTGTTT ATTCTAACTT TTATGATATA CTGTTTTTCT GACAGATTTA TCAAAACAAT	3240
TACTTGAAAA AGCTCATGGT GGGTTAAAA TAAATCCGGA TGAGCAAAGA CGCTATCTTG	3300
GTACTTTTGA GGAAAGAGTT CTTGGATATG TAGATATTGA CACAGCAAAT AGCCCTCAGT	3360
TAGAAAAAGG CTTTTTATTT ATTTTAGAAA ACCTTCAGGA AAAAGCAGAG CCACTATTTG	3420
TGAAGATTTT ACCAACTATC GAATTTGATA AGCAAGTTTT CTACTTAAAA GAAGCAAAAG	3480
AAACTGATAG TCAAGCCACC ATAGTATCTG AAGAGCATAT TACTTCTCCT TTTGGCCTGG	3540
TTATTATAG CAATGCACCA GTTCAAGTAG AAGAAAAAGA CCTTCGACTT GCTTTTCCAA	3600
AACTTTGGGA AGTTAAAAAG GAAGAACCAG CCAAAACATC CTTATGGAAG AAATGGTTTA	3660

GCTAAATCIT GCACATATTT AATAAGTGCC CAATATTGGC AGCCGTGCGC TCCAGATAGA	3720
AAC TGGCATT TTTCAAATA TCTTCTAAAG GTTCACTTTT CTCCAAAATA GAAAAGACAG	3780
CTTGGATATT TTCAAATGGT AGGGGAGGTA AATCTTCAGC AAGACTACCG CAAATAGCAA	3840
TAACAGGAAC TCCAACAGGG GTTCTTTTGT CAACACCTAT AGGCGCTTTC CCAGCAAAGC	3900
TTTGAATATC AAGTCTTCCT TCTCCAACAA CAACCAAGTC AGCATCTGAA ACTTTCTTAT	3960
CAAAGTTGAT TAAGTCCAAG CAGGTATCAA TTCCAGACAC GATACTTGCC TGAGCAAAGG	4020
CACACAAACC ACCAGCAAGG CCTCCACCTG CTCCTGCTCC TTTAATTTCT AATGTTGCAG	4080
GTGAGAATTT TTCATAAAAA TCTTGGATCG CCTGATCTAC GACTGCAAAC ATAGTCGGAT	4140
GTAGACCTTT TTGATTGCCA AAAGTGTAAG TCGCACCTTG ATGACCACAT AAGGGACTCA	4200
CGACATCTGC TAAAATATGA ATTTGAACAC CTTCAGGAAT TTTATAGCAA TTTTCTGTG	4260
AAACAGAAGC TAAGTTAAT AAGGATTGAC CGGAAGCAGG CAAGACATTT CCATCCCTAT	4320
CATAAAATG ATAACCTAAA CCAGCAGCAA TCCCAGTCC TCCATCATTA CTGGCCGTGC	4380
CACCAACACC GATATAAATA TCTTTAATCC CTTTAGAGAT GAGATGAAGA ATCAACTCTC	4440
CAATACCACA AGTTTGGATT TGAAGTGGAT TTCGTTTCTC TAGCGGAATT TTTCCAAGAC	4500
CAACCAAGTC AGTACTTCA AATAGTGCCA GTTCCCCTTT TTGAAAATAG CGCATGGCTT	4560
CTTTTGTCC AAAAGGGTCT GTCACTTGA TCCATTTTTC TTTTAGGTCA AGAGAATGTC	4620
GGATAGCATC TACAGTACCT TCTCCCCAT CACCAACAGG GCAGAGGAGA CATTCTACAT	4680
CTGCTATCGA TTGTTGGAAG CCTCTTTTTA TTGCTTCAGC TACCTGTGA GCTGTCAAGC	4740
TTTCCTTAAA CGAATCCGGT GCAATTACAA TCTTCATATT TTCCCTCATT CTAAACAGTC	4800
AATCAAAGGG AGAAGTCTA AAAAATCCCT CTTGTCAACA TGATGTGGTA TTTCTTTTTT	4860
GAGCACTTCT TTGGCACAAA AGGCGATTCC TAACTTCGCC GACTTCAACA TTAATAGATT	4920
ATTAACCCCA TCACCGATTG CCACCGTTCT TTCTTTAGAA AGTTTGTAGT TCTTTCTCCA	4980
TTTTTCCAGA GTCTCTTTTT TGACCTGGGG ACTTATAATT TGTCCAATA ATTTTCTGT	5040
TAAAAGACCT TCTTTGACTT CAAGCTAGTT GGCAGTGAAA TAGGCAATAC CAAGGGATT	5100
TGCTAATCTC TCCAATATT GGTGTAAATC CACCAGACAC CAGACCAACT AGGATGCCAT	5160
TCTTTTGGAG AATAGAGATG AACTCTGGGA CATTTAGCGA TAGATGAATT GAGTTGAAGA	5220
CGTTATCAAA GACCAAATA GGAAGACCTT CCAACAAGGA CACTCTTTTT CTTAACTGC	5280
TTTCAAAGAC CAACTCTCCT CGCATTGCTC GACTTGTAAT CTGCGAAATT TCCGCCTCAT	5340
GACCTGCCTC TCTCCCTAAA AGATCAATCA CTTCTTCTAG GATTAAGGTT CCATCTACAT	5400

420

CCAAAAACACA CAAGCCTTTT ACTTGAGACA TCAGTCTCTCC TCTCTAAACA GCCTAAAAAT 5460
CGTATGAAGT CATCATACGA TTTTATCTAT TAATTAACTA AACTATGGTA CAAGTCAAGG 5520
TATGACTTGC AGGCTGTATC CCATGAGAAG TCACTCTCCA TAGCTTGT TTGTAGGTTT 5580
CTCCAAATGT CTGGATGGT TCTATACAAG TCCAATGCTG TTTGGAAAGT CCAATTTAAC 5640
CAATAAGGAG ATAGATTGTC AAAGCTAAAG CCAGTACCGC TTCTTCGAT TGGATTGAAA 5700
GCGCGAACTG TATCTCGCAA GCCTCCAAC TCAATGGACCA ATGGCAAGGT TCCATAACGC 5760
ATAGCCATCA TTTGAGACAA GCCACACGGT TCAAAACGAC TTGGCATGAG GAAGAGGTCA 5820
CAAGCAGCGT AGATTTCTTG AGCAAGTTTG ACATCAAAAG TGATATTTGT TGATAGCTTG 5880
TCTGGGTAAA TCTGAGCAAA CCATGAGAAA GCTCCTTCAA AGGCTGGATC GCCAGTTCCC 5940
AAAAGAACAA TCTGAACATC TTCTTGCAAG ATATGGTGAA GACTTTTCGAC CACCACATCA 6000
AAACCTTTTT GACGTGTCAA ACGAGAAACA ATTCCACCA GTGGAACGTC TGCTCTAACA 6060
GGCAAGCCAA CTCTTTCTTG CAATTTTGCC TTATTTTGG CTCTCCAGA CAAATCTTCC 6120
TGATTGAAAT GATAGTCTAA AAGAGCATCC GTCTGAGGAT TATAAAGATC AGCATCAATC 6180
CCATTCACGA TACCAGATAC TTTACCAGAC TCCATTTTAA GAATCTGATC CAAATTACAT 6240
CCAACTGAC TAGTCATAAT TTCATGAGCA TAGCTAGGTG AAACGGTTGA AACACGGTTC 6300
GCATAGAGAA TACCTGCCTT CATCCAGTTC AGACAGTGT TCCATCGAAG GGTGCCATCA 6360
GCGTAACGTT CAAAGCCAAC TCCAAACAAA TCACCCAACA TTCCTTCTGA AAATTGTCCT 6420
TGGAATTCTA AATTATGAAT GGTAAAACT GTTTCAATGT CCTCATAGGC TTGAATCCAA 6480
CGGTATTTTT CCTTCAACAA GAAAGGAATC ATAGCTGTAT GGTAGTCATG AACATGGAGA 6540
AGATCAGGAA TAAAGTCAAT CCTTTCCATA GCCTCAATGG CAGCCAGTTG GAAAAAGGCA 6600
AAGCGTTCTC CGTCATCAAA ATCACCGTAA ACATGACCAC GGAAGAAATA ATATTGATTG 6660
TCAATAAAGT AGAAGGTAC ACCATTTAAT ACTGTTTTCT TAATTCCACA ATACTGTCTG 6720
CGCCAACCAA CGCTCACCTC AAAATGAAGC ACATCTTCAA TCTGATTTC AAATTTAGCC 6780
TCTACCATAT CATAGTAGGG TAAATCACT GCAACTTCGT GCCCAGCTTT TACCAGTGAT 6840
TTTGGAAGAG CGCCAATGAC GTCTCCCAA CCACCTGTTT TTGAAAAGGG TGCACCCTCT 6900
GCTGCTACAA ATAAATTTT CATGAATGAA TATCCTCTGT TACTTTAGCA CCTTTCTTAA 6960
CCACAACCTG ATGTTCTGCA GTTCTCGAA TCACAACACC ATGCTCAACT TCAACCCTT 7020
TGTCCAAGAT AGCATATTCG ACCTGAGCCC CTCTCCAAT AACAAACGA GGAAGAGCA 7080
GGCTATCTTT AACCAAGCTA TCCTTATGGA CATGAATATT ACGTGATAGA ACAGAATTAG 7140
CTACTTGACC TTCAATAATA CTACCAGAGG CAACTGAGA AGTGCTTACC TTAGATGTAT 7200

421

TAGCATAGTA AGTTGGCTCT TCGTTTTTGA CCTTTGTATA AATCTTTTGG TTTGGTGAGA	7260
AAAGAGAATA GAATTTTTGT GATTCAAGCA TATCGATATT CGCTTGATAA TAAGATTTAA	7320
CAGAGTGAAT ATTGGCTAGA TAGCCCGTGT ACTCGTAGGC GAAAGCTCCC TCTTTTACAG	7380
CCAAATCCCG TAAACATAG CGCAATTTCT CTGGATGTTT TTTTITAGCT TCTTCTTCCA	7440
AGTGTTCAAT CAACCAAGGT GTATCAACGA CAAAGATATC TGTAGACATA TTGAACGTTT	7500
CAGCTGTTGA CTTGCTATCA AAGAGTTTAT GAGAAAGAAC ATGGTCTGTT TCATCTACAT	7560
CCAAGATTGC ATTTACTTCT GAAATATCTT TCTTAGCTAG TTTTITATAA ACTACAGTGA	7620
TAGGCTCTTT TGTGTACTA TGTAGGTGGA AAAGTTGGT CAAATCAATG TTAATAAGAA	7680
CATCGCAGTT GAGGGCAACC GTTTGGTTTG AGCCAGAACG TTTCAAATAA GTAAGAAGCT	7740
GTTGGTAGTA TTCTTTTCCA ACTGTACTAC TTTCTACACG GGTATTGTAA ATTCCTAGAT	7800
AGTAATGGCT AAGAAGGGTT GATAAGCCCC ACTCGCGTCC TGAACGAATA TGGTCAAATA	7860
CTGAGCTGAT ATTATCCTGC TGGAAAATAC CAAAGACACT ACGAACACCT GCATTAGCAA	7920
GGCTTGAAAG TGGGAAGTCA ATCAAACGAT ATTTCCACC AAATGGCAAA CTGCTACTG	7980
GACGGTGGTC CGTCAATGTC GACATATTGT GAAAACCAAC TGTATTTCCT AAAATGGCAG	8040
AATATTTATC AATCTTCATC TGTGCTACC CCCACTACTT CATTTATATCC TACAACTTGT	8100
ACTTCATCTG TTCCATCAAT TTCGACACCG TCAGAAATAA TCGCACCTTC ACCAATAATG	8160
GCACGTTTAA TCTTAGCTCC TTGACCAATG ATAGCTCCAC TCATGATAAC TGAATCAAGG	8220
ACTTCCGCTC CTTGCGGAAC TTGCGCGCCT GTTGAAAGGA TAGAATGTTT AACAGTTCCA	8280
TCAACGAAAC ATCCGTCTAC AACTAATGAG TCTTCCACAT GAGCATTTGC CCCGAGGAAG	8340
TTTGGTGGTG AAATCAAGTT TCTTGAGTAA ATCTTCCATT GACGGTTACG ACTATCCAAG	8400
GCATTTTCTG GAGAAATATA CTCCATGTTT GCTTCCCAA GTGACTCAAT AGTACCAACA	8460
TCTTTCCAAT AACCCTAAA TTCGTAAGCA TAAACACTTT CACCTGACTC AAGGTAATTT	8520
GGAATGACAT TTTTACCAA GTCTGACATG CCAACCTTGC TCTTTTCAGC AGCGACTAAC	8580
ATATTACGAA GCGGTTGCCA ATCAAAAATG TAGATTCCCA TAGAAGCTTT TGTAGATTTA	8640
GGTTGAGCTG GTTTTCTTTC AAATCAACA ATACGATTGT TAGCATCTGT GTTCATGATA	8700
CCAAAACGGC TTGCTTCTTT AAGAGGGACG TCTAAACTG CTACTGTCAA GCTGGCATT	8760
TTATCCTTAT GAGACTGGAG CATATCATCA TAGTCCATTT TGTAGATGTG ATCCCAGAC	8820
AAAATCAAGA CATACTCAGG ATTGACACTG TCGATATAGT CGATATTTTG GTAAATAGCG	8880
TGACTAGTCC CCTCAAACCA ACGATTTTCCT TCACTTGACG AATAAGGTTG AAGAATAGAG	8940

422

ACACCTGAAT TAATACCGTC TAGTCCCCAG CTTGAACCAT TCCCAATATG GTTGTGAGA 9000
GCAAGTGGTT GATACTGTGT AACGACCCCA ACATTGTGAA TCCCTGAGTT GGCACAGTTT 9060
GATAGGGCAA AGTCAATGAT ACGGTAGCGC CCACCAAATT GCACAGCTGG TTTTGCGATG 9120
CTTTGAGTGA GTTTACCGAG ACGAGTTCCT TGCCCACCAG CAAGAATCAA AGCTAACATT 9180
TCATTTTCA TTTTCTACTC CTTTGTGGT TTTATTTGTG ACGGTTTTAG TAGATTTCAA 9240
GCGACGTTTG ATTTTCCATA CACTTGCTCC CATAGCCGGT AGGGTAAAGG TTAAGGTCTG 9300
CTCATAATCT TTCCATAGTC CTTCTTGGT TTGAACAGTT TGATTATGTT CTTTCCAAAC 9360
GCCTCCCCAC TCTTCCAACT CAGTATTTCA TACTTCTCG TAAATTCCTG CAACGGGTAG 9420
TCCGATTGTA AAATCTTTCC GCTCAACAGG TACCATATTA AAGATACAGA CTAACATTTT 9480
TCCCTTTTCA CCCTTACGAA TAAAGGAAAG AACACTCTGG TCTCGATTAT CCGCATCAAT 9540
GATTTCATA CCATCATAGC TGGTATCAAT TTCCCACAGA CAGCGATGAT CTTTGTAATA 9600
CTGGTTTAGC TGAGAAGCGA AATACTTCAT CTTAGCATTC ATTGGGTCTT CTAGGTTAGA 9660
CCATTCCAAC TGTCTTCAG ATTTCCATTC TAGGAATTGA CCGTATTCGC TACCCATGAA 9720
GAGCAATTTT TTACCAGGGT GACAAATTTG GTACGTATAG AGATTGCGCA AGCCTGCGAA 9780
TTGATTGTAA CGATCTCCCC ACATCTTATG CATCATACTC TTCTTGCCAT GAACCACTTC 9840
ATCGTGGGAG AATGGCAAGA GATAATTCTC CTTGAAAACA TACATAAAGC TGAAAGTCAC 9900
CAGGTAAAG TCATATTTAC GATAGATCGG ATCTTCTCG TAGAAACGGA GGATATCATT 9960
CATCCAGCCC ATGTTCATT TGTAGTCAA TCCTAGACCA CCAATCTCTT TCATTCCCGT 10020
AATCTTGATC GCAGACGAAC TTTCTTCTGC AATCATCATC ACATCTGGAT ATTCTAACTT 10080
AATAACCTCA TTCAAGCGCT GAAGGAAATA ATAACCTTCA TAGTTGAGAT TTCCGCCATC 10140
TTTATTAGGT GTCCATGGAG CATCATCATA GTCCAAATAG AGCATGTTGC TAACAGCATC 10200
CACACGAATA CCATCCAAAT GATAGACATC AATCCAATGC TTAATGCAAG AAATTAAGAA 10260
GGACTGGACT TCATTTTTTC CAAGGTCAA ATTAAGGGCA CCCCAACCAT GGTTATGAGC 10320
CTTATTATGG TCTTGGTATT CAAAAGTCGG TGTCCCATCA TAATAGGCTA AGGCATCATC 10380
GTTGATGGTA AAGTGAAGTGG TACCCAGTCC ACAATAACCC CAATATTATG GGTATGACAC 10440
TCCTCGACAA AATCTTGAAA CTCCTCTGGT CGGCCATAAG CATGCTCTAA AGCGAAGTAA 10500
CCCATAAGCT GATACCCCA ACTCAAGCCC AAAGGATGGG ACATCAAGGG CATAAACTCA 10560
ATATGAGTAT AGTTCATTTT AACGAGATAA GGAATGAGTT CATCCTTGAG CTGGGCAAAA 10620
CTATAAGGAC TGCCATCAGA ATTTCTTTTC CATGATCCAG CGTGAAGTTC ATAAATATTG 10680
ACAGGACGCT CTTCAAAGCC CCAACGTTTT CTTCTGTTCA GCCAAAGTCC ATCCTTCCAT 10740

423

TTCTTCTCAG GAAGCTCTGT TACGATTGCC CCTGTTCCCTG GACGAGCCTC ATACCTGACA 10800
GCAAAAGGGT CAATCTTCAT CAGTTGATGA CCATTTTGAC GTGTGACATG ATATTTGTAA 10860
ATATGCCCTT CTGAGCCAT ATTGGTAAAG ACTTCCCAGA CCCCCAAATC ATTTCTTACC 10920
ATTGGAATCT GATTTTCAAT CCAGTTGGTA AAATCACCAA CCAAGTGAAC AGCCTGAGCA 10980
TTAGGTGCCC AAACACGGAA GGTATAGCCA TGCTCTCCAT TTAGTTCTTC CCTATGTGCT 11040
CCTAGATAAT GTTGAGATA AAAATTTTCA CCCGTCATAA AGGTTTMTAA TGCTTCTCTA 11100
TTATCCATAT ACTCCCCTTC TCCTGTAAGC GTTTTCTATG TTTTATTAT ACTACCTTTT 11160
TAGAGAAGAT TCAAGTAAAT TACTATACTT CTTTAATTAT TTTGAAAATC TACAACAAGT 11220
TCACTTACTC GTTCAATTGT AAATCAATAT TTTTCAAAA AATTGCGAAA ACGCCTTTCT 11280
TTTTCTACTA TAGTGAAATG AAATAAAACA TGCACAAATC GATTAAGGAA TTAAATCTAA 11340
TTTCTAACAA TGTCTTAGAA ATCAAAGTG ACTATTTTAA CTCC 11384

(2) INFORMATION FOR SEQ ID NO: 46:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7577 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TGTTGATTG TTAGTAGACG TTGACCAACG TCCTTCGGCT GGAAAAGGAA TTCTCCTTAG 60
TTTCCAACAC GTTTTCGCCA TGTTTGGTGC GACCATCTTG GTACCATGA TTTTGGAAT 120
GCCTGTATCT GTTGCCCTTT TTGCTTCAGG TGTTGGAACA CTCATCTACA TGATTGCTAC 180
TGGTTTTAAA GTTCCAGTTT ATCTAGGTTT TTCATTTGCC TTTATCACAG CTATGTCACT 240
GGCTATGAAA GAAATGGGGG GGGATGTATC TGCTGCCAA ACAGGGGTTA TCTTGACTGG 300
TTTGGTCTAT GTCCTTGTG CTACCAGCAT CCGATTTGTA GGAACAAAAT GGATTGATAA 360
ACTCTTGCCA CCAATCATT TCGGTCCTAT GATCATCGTT ATCGGTCTTG GACTTGCAGG 420
TTCAGCTGTT ACCAATGCAG GTCTTGTAGC AGACGGAAAT TGGAAAAATG CTCTGGTAGC 480
CGTTGTACT TTCTTAATTG CTGCCTTTAT CAATACAAAA GGAAAAGGCT TCCTACGAAT 540
CATTCCATT CTCTTTGCCA TTATCGGTGG TTACCTTTTC GCACTAACTC TTGGCTTGGT 600
TGACTTTACA CCAGTTCTTA AAGCCAACTG GTTCGAAATT CCTGGTTTCT ACTTGCCATT 660
TAGCACAGGT GGTGCCTTTA AAGAGTACAA TCTTTACTTT GTCCAGAAG CCATCGCTAT 720

424

CTTGCCAATC GCTATCGTAA CAATTTCTGA ACATATCGGA GACCATACTG TTTTGGGTCA	780
AATCTGTGGT CGTCAATTCT TAAAAGAACC AGGTCTTCAC CGTACTCTTC TTGGTGACGG	840
TATCGCAACT TCTGTTTCTG CCTTCCTTGG TGGACCAGCC AATACAACCTT ACGGAGAAAA	900
TACAGGGGTT ATCGGTATGA CTCGTATCGC TTCTGTCTCA GTTATCCGTA ACGCTGCCTT	960
CATCGCGATT GCCCTCAGCT TCCTTGGTAA ATTCACTGCC TTGATTTCOA CTATTCCAAA	1020
CGCTGTACTT GGTGGTATGT CAATCCTTCT CTATGGGGTT ATCGCCAGCA ATGGTTTGAA	1080
AGTCTTGATT AAAGAACGTG TTGATTTCGC TCAAATGCGA AACCTCATCA TCGCAAGTGC	1140
TATGTTGGTT CTGGACTTG GAGGAGCTAT CCTTAAACTT GGTCCAGTTA CACTTTCAGG	1200
TACTGCCCTT TCAGCCATGA CAGGAATCAT CTTGAACTTG ATCTTGCCAT ACGAAAATAA	1260
AGACTAAGAG TCTAAATACA CCTAATCCAC TCAGACAGCT GAGTGGATTT TTCGTATACC	1320
ATAATAAAAG TGTCTTAACA AAATTATTAA AATCAAAAAA CGTATAATAT CAGATATTCT	1380
AAAACCTTGA TACTGTACGT TTTATCATAG AAATTTTAC TTTATTTTCT CATCAAATGA	1440
GATTTGCATC AATCTCTGT CTTACTTGC TTTCTTCTTC GCTTTCTTCA TTTGTTAGC	1500
CATACGTTTC ATGGACTGTT TCATGGCAAA TTCACCAATT TTACCTTTCA AACCGCCACC	1560
AAACATCTGG CTCATATCTG GCATTCTGC TCCTCCGAGA GCTGATAAGT CAGGCATACC	1620
GCCTTGTCCT ATCATTCCTT CAAGGGCAGA CATATCCATT CCTCCCATAT TTGGCATATT	1680
TTTAGGAAGG TTATTTGGAT TAATCCCAT TTGCTTCATC ATTTTATTCA TATCCCCAGA	1740
CATAACACCC TGCATGAGCT GTTTAGCCTG GTTAAAGTCC TTGATGAATT TATTGACTTC	1800
GACGAATGTA TTTCCAGAAC CAGCAGCAAT ACGACGGCGA CGGCTTGGAT TTAACAAATC	1860
TGGGTTTTCA CGCTCTTCAG GTGTATCGA AGACACAATG GCACGTTTAC GAGCAATCTG	1920
GCGTTCATCC ACCTTCATGT TTTGAAGGGC TGGATTGTTG GCCATACCTG GAATCATCTT	1980
GAGCAAGTCT TCCATCGGCC CCATATTTTG CACCTGATCT AATTGATCGA TGAAATCATT	2040
AAAATCAAAG GTGTTTTCGC GCATCTTCTC AGCCATTTC AAGGCTTTTT GTTCATCGTA	2100
TTCTTGAGAA GCTTTCTCAA TCAAAGTGAG CATATCCCC ATACCAAGGA TACGGCTAGA	2160
CATGCGTCT GGGTGAAGG TTTCAATGTC CGTAATCTTT TCACCTGTAC CAGTGAACCT	2220
GATTGGTTTT CCAGTAATGT GACGAACAGA CAGAGCAGCA CCACCACGAG TATCGCCATC	2280
AATCTTGGTA AGGATGACCC CAGTCACTTC CAACTGAGCA TTAAACTCAC GCGCAACATT	2340
GGCTGCTTCC TGACCAATCA TAGCATCAAC GACAAGCAAG ATTTCAATTG GTTGAGCCAA	2400
TGCTTTCACA TCACGAAGCT CATTCATGAG GAGCTCATCA ATCTGCAAAC GACCCGAGT	2460
ATCAATCAAG ACATAGTCGT TATGATTAGT TTGGGCTTGC TCCAAACCTT GACGTACAAT	2520

425

CTCAACAGCT GGTACTTCTG TTCCAAGTGC AAAGACAGGC ACATCAATCT GTTGTCCTCAA	2580
GGTCTTAAGC TGGTCAATGG CAGCTGGACG ATAAATATCC GCCGCAATCA TCAAAGGACG	2640
AGCATTMTCT TCTTTCTTGA GTTTGTTGGC CAATTTACCA GCAAAGGTTG TTTTACCAGC	2700
CCCTTGTAAG CCAACCATCA TGATGATGGT TGGAATCTTA GGTGACTTGA TAATTTCTGC	2760
CGTATCAGAA CCTAAAACGG CTGTCAATTC CTCATCAACG ATTTTAATAA TCTGTTGCGC	2820
AGGATTAAAGT GTATCAATGA CCTCATGCCC GACTGCACGC TCACGAACTT TCTTGATAAA	2880
GTCCTTTTACA ACAGGCAAGG CAACGTCGGC CTCGAGCAAG GCCAAGCGAA TTTCTTTGGT	2940
TGCCTCTTGG ACATCAGATT CAGAGATTTT TCCTTTTTTA CGTAGATTTT TAAAGACGTT	3000
CTGCAACGCT TCTGTAAAC TTTCAAATGC CATTTTCTT CCTCTTATTC TCTATTATCA	3060
ATGCTTGTTA AAATTTCTAT CTGCTCCTGC AGAAAGTCAT CCTTGGGATA GCGCTCCAAA	3120
ATCTGATCAA AAATCTGACT GCGGACAATA TAGTCCGAGT ACATGTGCAA TTTTCATCTCA	3180
TAATCTTCCA GAATCTTTC TGTTCGCTTG ATATGTTCAT AGACAGCCTG ACGACTGACA	3240
CCGAACCTCT CGGCAATTTC AGCAAGGCTG TAATCATCAG CGTAGTAGAG CTCGATATAA	3300
TTCAATTGCT TATCTGTCAA AAGCGCCGCA TAAAATTCAA AGAGCGCATT CATACGATTG	3360
GTTTTTTTGA TTTCCATAAC TTTTATTATA CCAAAAATTA GCCTAATCTA CCACACTAGG	3420
AAGCCGATCC AAGAAGATAG ATAGCTAAAT TTGAAAAAGA CATGAGCCTA GCCCCAAGTA	3480
ATTTCCAATT GATAGCTGGC AAAGGGATGT CCTCTTGAT TTTGTAGTTG ATAATCTAGT	3540
TCAATCTTTT GCCTATCAAC TTGATAATGG CTCGTTTGGG TGATAAACTC CTGCATGCCC	3600
ATAGGTGTAG GAATATAGGC TAAACTATCG CTATCCTTTA GAAAGCGCAT AATGGTCTTG	3660
GGATTAGAAA ATCGGCTCAT CACAAGTTCT TGACCATGAA ATTTAATCAC TACTTTTTC	3720
TTTTCCTCAT TATAGAAAAG CAGGTAGCTA TAATCTCCTT TTTTCATGCAC TTCCACATCA	3780
TAAAGCTGGT CAATCACTTC CAACTGCTCA TCAAACTGAA TCGTATTTTCG CATCCGAATC	3840
TTACATCAG GCCCTCTTTC TTGTCTCTTG TCCTACTATT TTACCAAAAA GAGCAGGATT	3900
TTGCTATAAT GGTATATGA ACGAAAAAGT ATTCCGTGAC CCTGTTTACA ACTACATCCA	3960
TGTCAATAAT CAAATCATCT ATGACTTGAT TAATACAAA GAATTTTTCAGC GTTTGCGCCG	4020
GATCAACAA CTGGGAACTT CCAGTTATAC CTTCACGGT GGAGAACACA GTCGCTTCTC	4080
TCATGTCTA GGAGTCTATG AAATTGCACG ACGCATCACA GAGATTTTCG AAGAAAAATA	4140
TCCTGAGGAA TGGAATCCTG CCGAGTCTCT CTTGACCATG ACCGCTGCTC TCCTACACGA	4200
CCTTGGGCAT GGTGCCTACT CCCATACTTT TGAACATCTC TTTGATACAG ACCATGAAGC	4260

426

CATTACTCAG GAGATTATTC AAAATCCTGA GACAGAGATT CACCAAGTCC TGCTACAAGT	4320
GGCACCTGAT TTCCCAGAAA AGGTGGCCAG TGTCATTGAC CATACCTATC CTAATAAGCA	4380
GGTCGTGCAG CTCATTTCTA GTCAGATTGA CGCAGATCGC ATGGACTATC TCTTGCGCGA	4440
CTCCTATTTT ACAGGAGCAT CCTATGGGGA ATTTGACCTG ACTCGAATCC TCCGAGTCAT	4500
TCGTCCTATC GAAAATGGTA TCGCCTTTCA GCGCAATGGC ATGCACGCCA TCGAAGACTA	4560
CGTCCTCAGT CGCTACCAGA TGTACATGCA GGTTTATTTT CACCCCGCAA CACGCGCCAT	4620
GGAAGTTCTC CTACAGAATC TTCTCAAACG CGCCAAGGAA CTCTATCCTG AGGACAAGGA	4680
TTTCTTTGCC CGAACTTCTC CACACCTCCT GCCTTTCTTC GAAAAAATG TGACCTTGAC	4740
TGACTATCTG GCTCTGGATG ATGGCGTGAT GAATACCTAC TTCCAGCTTT GGATGACCAG	4800
TCCTGACAAG ATTCTTGAGC ATTTATCGCA TCGCTTTGTC AACCAGCAAGG TCTTTAAATC	4860
CATTACCTTT TCACAAGAGG ACCAAGATCA ACTTACTAGC ATGAGAAAAT TGGTTGAGGA	4920
TATCGGCTTT GATCCCGACT ACTACACTGC CATTACATAAG AACTTTGACC TCCCTTATGA	4980
TATCTATCGT CCCGAATCTG AAAACCCACG GACACAGATT GAGATTTTAC AAAAAAATGG	5040
AGAACTGGCC GAACTCTCTA GCCTGTCTCC TATCGTCCAA TCCCTTGCTG GCAGTCGCCA	5100
CGGAGATAAT CGCTTTTATT TTCCAAAAGA AATGTTGGAC CAAAACAGCA TCTTTGCAAG	5160
CATTACCCAG CAATTTTATC ACTTGATTGA GAACGATCAT TTTACCCCAA ATAAAACTA	5220
GAAGAGGAAA TTTATGAGTA TTAACTAAT TGCCGTTGAT ATCGACGGAA CCCTTGTCAG	5280
CAGCCAAAAG GAAATCACTC CTGAAGTTTT TTCTGCCATC CAAGATGCCA AAGAAGCTGG	5340
TGTCAAAGTC GTGATTGCAA CTGGCCGCCC TATCGCAGGC GTTGCCAAAC TTCTAGACGA	5400
CTTGCAAGTG AGAGACGAGG GGGACTATGT GGTAACTTC AACGGTGCCC TTGTCCAAGA	5460
AACTGCTACA GGACATGAGA TTATCAGCGA ATCCTTGACT TATGAGGATT ATCTAGATAT	5520
GGAAATTCCTC AGTCGCAAGC TCGGTGTCCA CATGCATGCC ATTACCAAGG ACGGTATCTA	5580
TACTGCAAAT CGCAATATCG GAAAATACAC TGTACACGAA TCAACCCTCG TCAGCATGCC	5640
TATCTTCTAC CGTACCCCTG AAGAAATGGC TGGCAAAGAA ATTGTTAAAT GTATGTTTAT	5700
CGATGAACCA GAAATTTCTG ATGCTGCGAT TGAAAAAATT CCAGCAGAAT TTTACGAGCG	5760
CTACTCCATC AACAAATCTG CTCCTTTCTA CCTCGAACTC CTTAAAAAGA ATGTAGACAA	5820
GGGTTACAGC ATTACTCACT TGGCTGAAAA ACTCGGATTG ACCAAAGATG AAACCATGGC	5880
AATCGGTGAT GAAGAAAAATG ACCGTGCCAT GCTGGAAATC GTTGGAAACC CCGTTGTCAT	5940
GGAAAAATGA AATCCAGAAA TCAAAAAAAT CGCCAAATAC ATCACCAGAA CAAATGACGA	6000
ATCCGGCGTT GCCCATGCCA TCCGAACATG GGTACTGTAA AAGTATCATT TTTCAATAAG	6060

427

AATTGATTAG CAATAAAATC CAATGAATTT TTTTAGCAAA CTATTTAATT TAAAACAAAA	6120
TAATCATAAT AGAGACACAA ATTCTGATTG TAACAATTTT TACCTAAACG AATTAGAATG	6180
TGGCCTTACT CCTGGGCAAC TCATACTCAT AGATTGGACT CAAAAACAG GGAGAAATTA	6240
TAATTTCCCA AGATATTTTA AATACTCTCT TCAAATTGAC CCTGAATCTA CACACAATCA	6300
ATTATACAAA TTAGGATACT TCACTAAAA TAAGACTTTA TCATATCTTA CAGTAGTAGA	6360
ATTAAAACT ATATTATCTA AACATAATTT AGCTACTTCT GGAAAAAAG CAGAATTAAT	6420
TACAAGAATA ATTAATAATG TTAACATTGA CAATTTAGAT ATTCCGTTTCG AATTTAACT	6480
AACAAAAGAA GCACAAAATC TTATTATCGA ACATAGTGAC TATATCAAAG CATACTATGA	6540
TAAAGACATA ACTATGGAAG ATTATTGTAA AGAAAAAAC AATATCTCTT TTAAGCAAC	6600
TTTGGTGAT ATAAATGGA GTCTCTTAA TAAACAAGCT CATAGGAATA CTGTATCAGG	6660
AGATTTTGGA TGCTTATCTA ACACACGAAA GGCTCAGGA AGACATTGG AACAAGAAGG	6720
TAATATTAAA CATGCTTTAA TATATTACAT AGAATCTTTC ATAATTACTA TTCAGGATT	6780
AGAAAACAA TTTTCAGCCA CTGATTATCC AGTATATTAT CCCGATTCCA TACCTGACTA	6840
CTCACTAAAA CATATTCAA CATTAATGGA ATCATTATCT GATGACGATT ATGATTTTGC	6900
TTTGTATGAA GCATTATTTT GCTTCTCAAT TTTGAATGCA AATCATTTT TATCTAAGGA	6960
AGATATTGAC TATTTAAGAG TTAATTTACC TCGTTCCACT GCTGAAGAAA TAAACAATTA	7020
CTTAAAGAAA TATGAATGTT ATAGTCCTTT AAATAATTTA GAACTTGACG ATTTTGAATA	7080
AATTGACTAT ACAAACATTT ATATACTCGA TATAGTCTCA ATTTTATCTG ATGATTGCCC	7140
AAATTTTCA ATAATAAAC GCATAATATT ATGGAGACAA TCCCCTATAT TATGCGTTCT	7200
TTTAATATCA AAGACTTTT GACAACTTC TTTGATATCT AATTACATGC CCCCTGCAGG	7260
AATCGAACCT GCAACTACTC CTTAGGAGGG AGTTGTTATA TCCATTGAAC TAAGGGAGCT	7320
AGATAAAAAC TCTGCTAAAT GAGCAGAGTT TTTTAGTCGA ATTAACGACG GATTTCCTTG	7380
ATACGAGCTG CTTTACCTTG AAGAGCACGC AAGTAGTACA ATTTTCGCACG ACGTACTTTA	7440
CCGTAACGAA CAACTTCGAT TTTTCAACA CGTGGAGTGT GGATTGGGAA GATACGCTCA	7500
ACACCTACAC CGTTAGAGAT TTTACGAACT GTGTAGTTT CTGAGATTCC AGCACCTTA	7560
CGTGCGATAA CAACACG	7577

(2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4945 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

428

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CCTCGCTGAT GATTGGTGCT GTTTTATTTG CTGGTCCAGC CTTGGCTGAA GAAACTGCAG	60
TTCTTGAAAA TAGCGGAnCT AATACAGAGC TTGTTTCAGG AGAGAGTGAG CATTTCGACCA	120
ATGAAGCTGA TAAGCAGAAAT GAAGGGGAAC ATGCTAGAGA AAACAAGCTA GAAAAGGCAG	180
AAGGAGTAGC GATAGCATCT GAAACTGCTT CGCCAGCAAG CAATGAAGCT GCAACTACTG	240
AAACTGCAGA AGCAGCTAGC GCAGCTAAAC CAGAGGAAAA AGCAAGTGAG GTGGTTGCAG	300
AAACACCATC TGCAGAAGCA AAACCTAAGT CTGACAAGGA AACAGAAGCA AAGCCCGAAG	360
CAACTAACCA AGGGGATGAG TCTAAACCAG CAGCAGAAGC TAATAAGACT GAAAAAGAAG	420
TCCAGCCAGA TGTCCCTAAA AATACAGAAA AAACATTAAA ACCAAAGGAA ATCAAATTTA	480
ATTCTTGGGA AGAATTGTTA AAATGGGAAC CAGGTGCTCG TGAAGATGAT GCTATTAAAC	540
GCGGATCTGT TGTCTCTGCT TCACGTCGGA CAGGTCATTT AGTCAATGAA AAAGCTAGCA	600
AGGAAGCAAA AGTTCAAGCC TTATCAAACA CCAATTCTAA AGCAAAAGAC CATGCTTCTG	660
TTGGTGGAGA AGAGTTCAAG GCCTATGCTT TTGACTATTG GCAATATCTA GATTCAATGG	720
TCTTCTGGGA AGGTCTCGTA CCAACTCCTG ACGTTATTGA TGCAGGTCAC CGTAACGGGG	780
TTCTGTATA CGGTACACTC TTCTTCAACT GGTCTAATAG TATTGCAGAT CAAGAAAGAT	840
TTGCTGAAGC TTTGAAGCAA GACGCAGATG GTAGCTTCCC AATTGCCCCG AAATTGGTAG	900
ACATGGCCAA GTATTATGGC TATGATGGCT ATTTTCATCA CCAAGAAACA ACTGGAGATT	960
TGGTTAAACC TCTTGGAGAA AAGATGCGCC AGTTTATGCT CTATAGCAAG GAATATGCTG	1020
CTAAGGTAAA CCATCCAATC AAGTATTCTT GTACGATGC CATGACCTAT AACTATGGAC	1080
GTTATCATCA AGATGGTTTG GGAGAATACA ACTACCAATT CATGCAACCA GAAGGAGATA	1140
AGGTTCCGGC AGATAACTTC TTTGCTAACT TTAAGTGGGA TAAGGCTAAA AATGATTACA	1200
CTATTGCAAC TGCCAAGTGG ATTGGTCGTA ATCCTTATGA TGTATTTGCA GGTTTGGAAT	1260
TGCAACAGGG TGGTTCCTAC AAGACAAAGG TTAAGTGGAA TGACATTTTA GACGAAAATG	1320
GGAAATTGCG CCTTCTCTT GGTTTATTTG CCCAGATAC CATTACAAGT TTAGGAAAAA	1380
CTGGTGAAGA TTATCATAAA AATGAAGATA TCTTCTTTAC AGGTTATCAA GGAGACCCTA	1440
CTGGCCAAAA ACCAGGTGAC AAAGATTGGT ATGGTATTGC TAACCTAGTT GCGGACCGTA	1500
CGCCAGCGGT AGGTAATACT TTTACTACTT CTTTAAATAC AGGTCATGGT AAAAAATGGT	1560
TCGTAGATGG TAAGGTTTCT AAGGATTCTG AGTGAATTA TCGTTCAGTA TCAGGTGTTT	1620

TTCCAACATG GCGCTGGTGG CAGACTTCAA CAGGGGAAAA ACTTCGTGCA GAATATGATT	1680
TTACAGATGC CTATAATGGC GGAAATTCCC TTAAATTCTC TGGTGATGTA GCCGGTAAGA	1740
CAGATCAGGA TGTGAGACTT TATTCTACTA AGTTAGAAGT AACTGAGAAG ACCAAACTTC	1800
GTGTTGCCCA CAAGGGAGGA AAAGGTTCTA AAGTTTATAT GGCATTCTCT ACAACTCCAG	1860
ACTACAAATT CGATGATGCA GATGCATGGA AAGAGCTAAC CCTTTCTGAC AACTGGACAA	1920
ATGAAGAATT TGATCTTAGC TCACTAGCGG GTAAAACCAT CTATGCAGTC AACTATTTT	1980
TCGAGCATGA AGGTGCTGTA AAAGATTATC AGTTTAACTT AGGACAATTA ACTATCTCGG	2040
ACAATCACCA AGAGCCACAA TCGCCGACAA GCTTTTCTGT AGTGAAACAA TCTCTTAAAA	2100
ATGCCCAAGA AGCGGAAGCA GTTGTGCAAT TTAAAGGCAA CAAGGATGCA GATTTCTATG	2160
AAGTTTATGA AAAAGATGGA GACAGCTGGA AATTACTAAC TGGCTCATCT TCTACAACTA	2220
TTTATCTACC AAAAGTTAGC CGCTCAGCAA GTGCTCAGGG TACAACCTCAA GAACTGAAGG	2280
TTGTAGCAGT CGGTAAAAAT GGAGTTCGTT CAGAAGCTGC AACCACAACC TTTGATTGGG	2340
GTATGACTGT AAAAGATACC AGCCTACCAA AACCCTAGC TGAAAATATC GTTCCAGGTG	2400
CAACAGTTAT TGATAGTACT TTCCCTAAGA CTGAAGGTGG AGAAGGTATT GAAGGTATGT	2460
TGAACGGTAC CATTACTAGC TTGTCAGATA AATGGTCTTC AGCTCAGTTG AGTGGTAGTG	2520
TGGATATTCG TTTGACCAAG CCACGTACCG TTGTTAGATG GGTCAATGGAT CATGCAGGAG	2580
CTGGTGGTGA GTCTGTAAAC GATGGCTTGA TGAACACTAA AGACTTTGAC CTTTATTATA	2640
AAGATGCAGA TGGTGAGTGG AAGCTAGCTA AGGAAGTCCG TGGTAACAAA GCACACGTGA	2700
CAGATATCAC TCTTGATAAA CCAATCACTG CTCAAGACTG GCGCTTGAAT GTTGTCACCTT	2760
CTGACAATGG AACTCCATGG AAGGCTATTC GTATCTATAA CTGGAAAATG TATGAAAAGC	2820
TTGATACTGA GAGTGTCAAT ATTCCGATGG CCAAGGCTGC AGCCCGTTCT CTAGGCAATA	2880
ACAAGGTACA AGTTGGCTTT GCAGATGTAC CGGCTGGAGC AACTATTACC GTTTATGATA	2940
ATCCAAATTC TCAAACCTCCG CTCGCAACCT TGAAGAGCGA AGTTGGAGGA GACCTAGCAA	3000
GTGCACCATT GGATTTGACA AATCAATCTG GTCTTCTTTA TTATCGTACC CAGTTGCCAG	3060
GCAAGGAAAT TAGTAATGTC CTAGCAGTTT CCGTTCCAAA AGATGACAGA AGAATCAAGT	3120
CAGTCAGCCT AGAAACAGGA CCTAAGAAAA CAAGCTACGC CGAAGGGGAG GATTTGGACC	3180
TTAGAGGTGG TGTTCCTCGA GTTCAGTATG AAGGAGGAAC TGAGGACGAA CTCATTCGCC	3240
TAACTCACGC AAGTGTATCA GTATCAGGTT TTGATACGCA TCATAAGGGA GAACAGAATC	3300
TTACTCTCCA ATATTTGGGA CAACCGGTAA ATGCTAATTT GTCAGTGAAT GTCAGTGGCC	3360

430

AAGACGAAGC AAGTCCGAAA ACTATTTTGG GAATTGAAGT AAGTCAGGAA CCGAAAAAAG	3420
ATTACCTAGT TGGTGATAGC TTAGACTTGT CTGAAGGACG CTTTGCAAGT GCTTATAGCA	3480
ATGACACCAT GGAAGAACAT TCCTTTACTG ATGAGGGAGT TGAAATTTCT GGTACGATG	3540
CTCAAAAGAC TGGTCGTCAA ACCTTGACGC TTCATTACCA AGGCCATGAA GTTAGCTTTG	3600
ATGTTTGGT ATCTCCAAAA GCAGCATTGA ACGATGAGTA CCTCAAACAA AAATTAGCAG	3660
AAGTTGAAGC TGCTAAGAAC AAGGTGGTCT ATAACCTTGC TTCATCAGAA GTAAAAGAAG	3720
CCTTCTTGAA AGCAATTGAA GCGGCCGAAC AAGTGTGAA AGACCATGAA ACTAGCACCC	3780
AAGATCAAGT CAATGACCGA CTTAATAAAT TGACAGAAGC TCATAAAGCT CTGAATGGTC	3840
AACAGAAATT TACGGAAGAA AAGACAGAGC TTGATCGCTT AACAGGTGAG GTTCAAGAAC	3900
TCTTGGCTGC CAAACCAAC CATCCTTCAG GTTCTGCCCT AGCTCCGCTT CTTGAGAAAA	3960
ACAAGGCCTT GGTGAAAAA GTAGATTGA GTCCAGAAGA GCTTACAACA GCGAAACAGA	4020
GTCTAAAAGA TCTGGTTGCT TTATTGAAAG AAGACAAGCC AGCAGTCTTT TCTGATAGTA	4080
AAACAGGTGT TGAAGTACAC TTCTCAAATA AAGAGAAGAC TGTATCAAG GGTTTGAAAG	4140
TAGAGCGTGT TCAAGCAAGT GCTGAAGAGA AGAAATACTT TGCTGGAGAA GATGCTCATG	4200
TCTTTGAAAT AGAAGGTTTG GATGAAAAAG GTCAAGATGT TGATCTCTCT TATGCTTCTA	4260
TTGTGAAAT CCCAATTGAA AAAGATAAGA AAGTTAAGAA AGTATTTTTC TTACCTGAAG	4320
GCAAAGAGGC AGTAGAATTG GCTTTTGAAC AAACGGATAG TCATGTTATC TTTACAGCAC	4380
CTCACTTTAC TCATTATGCC TTTGTTTATG AATCTGCTGA AAAACCACAA CCTGCTAAAC	4440
CAGCACCAACA AAACACAGTC CTTCCAAAAC CTACTTATCA ACCGACTTCT GATCAACAAA	4500
AGGCTCCTAA ATTGGAAGTT CAAGAGGAAA AGGTGCGCTT TCATCGTCAA GAGCATGAAA	4560
ATACTGAGAT GCTAGTTGGG GAACAACGAG TCATCATACA GGGACGAGAT GGACTGTTAA	4620
GACATGTCTT TGAAGTTGAT GAAAACGGTC AGCGTCGTCT TCGTTCAACA GAAGTCATCC	4680
AAGAAGCGAT TCCAGAAATT GTTGAAATTG GAACAAAAGT AAAAACAGTA CCAGCAGTAG	4740
TAGCTACACA GGAAAAACCA GCTCAAAATA CAGCAGTTAA ATCAGAAGAA GCAAGCAAAC	4800
AATTGCCAAA TACAGGAACA GCTGATGCTA ATGAAGCCCT AATAGCAGGC TTAGCCAGCC	4860
TTGGTCTTGC TAGTTTAGCC TTGACCTGA GACGAAAAG AGAAGATAAA GATTAAATAT	4920
CGAAAAATCT TGTGAAATCT TTCCG	4945

(2) INFORMATION FOR SEQ ID NO: 48:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25002 base pairs
 (B) TYPE: nucleic acid

431

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

GACAACTCAA GTAGCTTTT CTTATTTTGA AAAAGGAGAT CAGAGTTTAA CTATGTCAGA	60
AAAATCACAA TGGGGGTCGA AACTTGGTPT TATTCTAGCA TCTGCTGGCT GGCCATCGGG	120
CTTGGTTCCG TTTGGAAGTT TCCCTACATG ACTGCTGCTA ATGGCGGTGG AGGCTTTTTA	180
CTAATCTTTC TCATTTCCAC TATTTTAATC GGTTCCTC TCCTGCTGGC TGAGTTTGCC	240
CTTGGCCGTA GTGCTGGCGT TTCCGCTATC AAAACCTTTG GAAACTGGG CAAGAATAAC	300
AAGTACAACT TTATCGGTG GATTGGCGCC TTGCCCCCT TTATCCTCTT ATCTTTTAC	360
AGTGTATATCG GAGGATGGAT TCTAGTCTAT CTAGGTATTG AGTTTGGGAA ATGTGTTCAA	420
CTTGGTGGA CCGGTGATTA TGCTCAGTTA TTTACTTCAA TCATTTCAA TCCAGCCATT	480
GCCCTAGGAG CTCAAGCGGC CTTTATCCTA TTGAATATCT TCATGTATC ACGTGGGGTT	540
CAAAAAGGA TTGAAAGAGC TTCGAAAGTC ATGATGCCCC TGCTCTTTAT CGTCTTTGTT	600
TTTATCATCG GTCGCTCTCT CAGTTTGCCA AATGCCATGG AAGGGGTTCT TTAATCTCTC	660
AAACCAGACT TTTCAAACT GACTAGCACT GGTCTCCTCT ATGCTCTGGG ACAATCTTTC	720
TTTGCCCTCT CACTAGGGGT TACAGTCATG TTGACCTATG CTTCTTACTT AGACAAGAAA	780
ACCAATCTAG TCCAGTCAGG AATCTCCATC GTAGCCATGA ATATCTCGAT ATCCATCATG	840
GCAGGTCTAG CCATTTTCCA AGCTCGATCC CCCTTCAATA TCCAGTCTGA AGGGGGACCC	900
AGCCTGCTCT TTATCGTCTT GCCTCAACTC TTTGACAAGA TGCCTTTTGG AACCATTTTC	960
TACGTCCTCT TCCTCTTGCT CTTCTTTT GCGACAGTCA CTTTCTCTGT CGTGATGCTG	1020
GAAATCAATG TAGACAATAT CACCAACCAG GATAACAGCA AACGTGCCAA ATGGAGTGTT	1080
ATTTTAGGAA TTTGACCTT TGCTTTGGC ATTCTTTCAG CCTATCTTA CGGTGTCATG	1140
GCGGATGTTC ACATTTTGG TAAGACCTT TTTGACGCTA TGGACTTCTT GGTTCCTAAT	1200
CTCCTCATGC CATTTGGAGC TCTCTACCTT TCACTTTTTA CAGGCTATAT CTTTAAAAAG	1260
GCTCTTGCAA TGGAGGAACT CCATCTCGAT GAAAGAGCAT GGAAACAAGG ACTGTTCCAA	1320
GTCTGGCTCT TCCTTCTTCG TTTCTTCGTT TCGTCATTCC AATCATCATC ATTGTGGTCT	1380
TCATTGCCCA ATTTATGTAA TCAAAAAGGA CTTGAGTAGT GAACTCAGGC CCTTCTTTT	1440
TATGGATGGC TAACAATCAA TTCCAAACCT TGCCCTTCCA GAGTCCAAGC TTCAACATCA	1500
CTTGGTAGGA TAAAGTGGCT GCCTTTTGA ATTGGATAAT TTTTCCCGTC AACAGTTAGC	1560

432

TGACCTTGAC CAGCCAAGAC ACTCAATAAG CTGTAGTCAG CTGTCTTTTC AAAGTCAACT	1620
TTTCCAGTAA TTTCCCACTT GTAAACTGCG AAGAAATCAT TAGATACAAG GAGAGTGGA	1680
CGCAAATCAT CTGCTTTAAC AGTTACAGGA CGGCTATTG CTGGCTCACC AATGTTCAAG	1740
ACATCGATGG ATTTTTC AAG ATGAAAGTTCA CGCAAGTTGC CTTTGTATC CTTGCGGTCA	1800
AAGTCATAGA CGCGATAGGT GGTATCGCTA GACTGCTGGG TTTCAAGGAT TAAGATACCC	1860
GCCCCGATAG CGTGATAGT CCCGCTTGGT ACATAGAAGA AATCTCCAGC CTTAACAGGG	1920
ACTTTGGTCA ACAAGTCATC CCAGTTCTTG TCCTCGATTT GCTGGCGGAG TTCTTCTTTT	1980
GACTTGGCAT TGTGACCGTA GATAATCTCT GAACCTTCAT CCGCTGCGAT AATGTACCAG	2040
CATTTCTGTTT TTCCGAGTTC GCCTTCATGC TCGAGTCCAT AAGCATCGTC TGGGTGAACT	2100
TGGACACTGA GCCAGTCGTT GGCATCGAGG ATCTTGGTCA AAAGTGAAA TACAGGTCTT	2160
GGACGATTGC CAAATAATTC ACGGTGTTCC GCATACAAAG TAGCAAGATC TGTTCCTCG	2220
TAACGACCAT TGGCAACTTT AGAGACTCCA TTTGGATGGG CTGAGATGGC CCAATATCT	2280
CCGATTTTCT CACTTGGGAT GTCGTAGCCA AACTCATCAC GTAGCTTGGC TCCACCCAG	2340
ATTTTTCTT GCATAACTGA TTGTAAAAAT AATGGTTCTG ACATGTCGAT CTCCTGTCTG	2400
ATTTTTCTCC CCTCATTATA GCAAAAAAG AGTTCGAATT GAACTCTTTT TTACATCTTA	2460
TAAAGCAGGG AGAAGATTTT ATAAAAATAG TAAACAAATG TGCTCTACCC GATGCTTGCA	2520
CCATTGCTAT AAATGACATC CTTGTACCAA TAGAAGGACT TCTTCTTGCT ACGTTTGAGA	2580
GCTCCGTTTC CTACATTATC TCGATCTACA TAGATAAAGC CATAGCGCTT ATTCAATTCC	2640
CCTGTGCCAG CTGAAACCGG ATCGATACAG CCCCAAGTCG TATAACCAAG CAAGTCAACC	2700
CCGTCTTGGT AAATGGCATC TCGCATGGCC TTGATGTGGG CCTCTAAGTA AGTAATCCGA	2760
TAGTCATCTG CTACATAACC ATTCTCATCC GGTGTATCCA TAGCACCGAG TCCATTTTCT	2820
ACGATAATAC TAAACTAAAA TCAAAAAGCA TTATATAATA GTGATATGAA ATCAACTAAA	2880
GAAGAAATCC AAACCATCAA AACACTTTTA AAAGACTCTC GTACAGCTAA ATATCATAAA	2940
CGCCTTCAAA TCCTTCTATA GTAAAATGAA ATAAGAACAG TACAAATCGA TCAGGACAGT	3000
CAAAATCGATT TCTAACAATG TTTTAGAAGT AGGGGTGTAC TATTTCTAGTT TCAATCTACT	3060
ATATTTCTGC TGATGGGCAA ATCTTATAAA GAGATTATAG AACTTTTATA GTAGTTTGAA	3120
ATAAGATGTG AACAACTCTA TCAGGAAAGT CAAATTAATT TATAGAAATA TTTTAGCAGC	3180
CAAGGTGTAC TGTATAGAT TCAATACACT ATAGACTGTA ATCAAACAAC GATTTGGCGA	3240
AATGTAAAAA AATATGAGGA GTTCGGACTC GACTCTCTCC TTCAAGAAAC ACGTGGTGGT	3300
CGTAACCATG CATATATGAC AGTTGAGGAA GAGAAAGCCT TTCTTGCCCG CCATTTGAAG	3360

433

GCTACAGAGG CAGGAGAATT TGTACAATT GATGCCTTAT TTCAGGCTTA TAAAAAGGAG	3420
TTAGGTCGTT CCTACACACG TGATGCCTTC TATCAACTGT TGAAGCGCCA TGGTTGGCGA	3480
AATATTACGC CACGTCCAGA ACATCCTAAG AAAGCAGACG CTCAAACCAT TGTTCGCTCT	3540
AAAAATAAAA TCTCAATCCA AGAAGGCAAG AAAGCGTTTT AAATATAGTA GACGTTTTCG	3600
TAAGGTTTGC TTGATGTACC AAGCTGAAGC TGGTTTCGGT AGAATCAGTA AACTGGGATC	3660
TTGTTGGGCT CCAATAGGAG TAGGTCCACA TATCCATAGT CACTATATAC GAGAATTTTCG	3720
CTATTGTTAT GGAGCTGTTG ATGCCTATAC AGGCGAATCA TTTTCTTAA TAGCTGGTAG	3780
ATGTAATACT GAGTGGATGA ACGCCTTTTT AGAAGAGCTT TCACAAGCTT ATCCTTTTAC	3840
TCGTTATGGA CAATGCTATA TGGCATAAAT CAAGTACCTT AAAGATTCCG ACTAATATTG	3900
GTTTTGCATT TATTCCTCCA TACACACCAG AGATGAACCC CATGGAACAA GTGTGGAAG	3960
AGATTTCGTAA ACGTGGATTT AAGAATAAAG CCTTTCGAAT TTTGGAAGAT GTCATGAATC	4020
AACTCCAAGA TGTCATACAA GGATTGGAGA AGGAGGTGAT AAAGTCCATC GTTAATCGGA	4080
GATGGACTAG AATGCTTTTT GAAAGCAGAT GAGTATTATA TGCAATTCTT TTATATAAAA	4140
AGACCGGATT GCTCCGATCT TTCAATAGTT CATATTCTCA ATTTCTATTT TAAAAATAGC	4200
TAAGGTTAAC GTCAAATGAC TACGCGACCT ATTTTCATACG ATAAAAATCA AGCACTAGAC	4260
CAGCAGGTCC TTGAACTAAT AAGGACTCTG TTCCCAATC GGTTACAGTT GGTCCGTGTA	4320
AAACCTTTAT ACCAAGCTCG TTCAACCGTT TGAGTTCTG GTCTACATCC TCAACCTCGA	4380
TATGAATAAT GATTCCTGAC TGAAAGTTTT CCAAAGGAAC CAAATGATTT TGTGACAACA	4440
TAAGGCAGTG ACTACCAATC GTAACTGAG CAAAACCATC ATTAGCATAA TCTGCCTTTT	4500
TATCCAAGAT ATGCTCCAAG TCAGCACAGA CTTGGGGAAC ATTTGAAACG ATAATATCTA	4560
ATTGATTTAA ATTCATTTAC TCTCCTCCAT AAAAAGACCG GATTGCTCCG ATCTTTTAAA	4620
GTTCTGCTCT ATGAAAATCA AAGAATAAAG TCTACAAGTT TCATATTTGA TTTTCGGCGA	4680
GAGGAATTAT TTAATGCGC GTGATTGCAA TCCTTCTTCT TCCAAGAAGA GACGGAATGG	4740
TACGAGTTCT TCTGCTTCGT ATTTTTCCTT GAAGGCTTG ATAGCTTCTT CTGAGTGAAG	4800
TTTTGGATCC AATTCAAGTA CTTCTACTGG AAGTGGACGG TGGTTGAGTA TGCGAGCATC	4860
GATGACAACA GTTTTACCTT CTTTGTTCAA TTAAACAGCT TCTGCAACAA CTGCATCGAT	4920
GTCTTCGATA CGGTCAACTG TGAATCCAAC AGCTCCTTGA GCTTCCGCAA TTTTAGCGTA	4980
GTCAGCGTTT GTGAAGTCTA CACCAAACAA GTGTTTGTIT GTATCTTCGT ATTTGTCTT	5040
GATGAAGCCG TACTCAGCAT TTGAGAAGAC AAGGTTGATA ACTGGAAGGT CGTATTGAAC	5100

434

GTTTGTGATA	ACGTCTGGGT	AGCACATGTT	GAATGCTCCG	TCACCCATGA	TGTTCCATAC	5160
TTGGCGATCT	GGATTGTCTT	TCTTAGCAGC	GATACCACCA	GGAAGGGCAA	TACCCATTGT	5220
CGCAAAGAGT	GGAGATGTAC	GCCACATGTT	CTTAGGTGTC	ATGTGAAGGT	GACGAGTAGA	5280
TGTTTGAGTA	GTGTTACCTA	CGTCGATTGA	GATAGATAGC	TCTTGATCAG	CATGTTTGTT	5340
GATTGCATTG	TAAACTTGAT	ACAATTGCAA	TTCACCCCTCA	GTTTTACCTT	CGAGTTTGTT	5400
CATGTAATCA	CGCCAGTTTT	GGTTGTTCTT	AACGTTTGCA	CGCCACCATG	GAGTTGATTC	5460
AACTGGGTTT	ACTTTGTCAA	GGATAGCTTT	AGCTGCTTGA	CCAGCATCAC	CAAGGATTGA	5520
AGCGTCAAGG	GCATGACGTT	TACCAAGTTT	GTAAGGGTCG	ATATCGACTT	GGATGAATTT	5580
TTCAGTGTTC	TTGAATGCTT	CGTAAACTTC	AGCAAATGGG	AAGTTTGAAC	CAAGGAAAAG	5640
AACGTGTGCT	GCTTCAAAGA	CCACTTCGTT	GGCTGGTTTC	CAACCAACAC	GGTAAGCAGA	5700
ACCTGTCAAA	CCTTCATAGT	TCCATTCGAA	AGCTTCAAAG	TTTTTACCAG	TTGTGATGAT	5760
TGGTGCTTTG	ATTTTACGTG	ACAATTCACT	AATCACTTCA	CCAGCTTTAA	CACCACCAAA	5820
TCCAGCATAG	ATAACTGGGC	GTTCAAGATT	TCAACAGCTT	TGTCGATTTC		5880
AAC TTCGTC	AAAGCAGGAG	CGATGAATGA	GCGTTCGTAT	GAACCTGAAC	CGTAGTATGA	5940
GTTTTTCATC	ATTTCTTGGA	AACCGAAGTT	TACTGGAATT	TCAACAACAG	CTGGACCTTT	6000
TTTAGAAACT	GCAGCACGGC	AGGCTTCGTC	AATTACTTTT	GGCAATGCT	CAGCGTAAGC	6060
TACACGTTTG	TTGTAAACAG	CGATACCGTT	GTACATTGGG	TTTGGTTAA	GCTCTTGAA	6120
AGCATCCATG	TTCAATTCGT	TAAC TGACG	TGATCCAAGG	ATCGCTAGGA	ATGGAGTGTT	6180
ATCCATAGCT	GCATCGTAAA	CACCGTTAAT	CAAGTGAGTC	GCACCTGGAC	CACCTGAACC	6240
AACTGCAACC	CCGATTGAGC	CGCCGAATTT	AGCTTGCTAT	ACCGCTGCAA	GAGCACCTGT	6300
CTCTTCGTGG	CGAACTTGTA	AGAAACGGAT	ATCTTTGTCT	TCAGCCAAAG	CGTCCATCAA	6360
TGAGCTGAGT	GTTCCTGATG	GGATACCGTA	GATTGTATCT	ACGCCCATG	TTTTCAATAC	6420
GTTAAGCATT	GCTGCAGATG	CAGTAATTTT	CCCTTGAGTC	ATAATGATAA	CTCTCCTTCA	6480
ATTTTTTTAA	ACTTGGAGAA	TACGATTACA	TAGAATTGGA	AACGTTCTCC	AAATTTTAC	6540
TATTCACCTG	TATCATATTT	ATGCTGACTT	TTCTAAAAAT	CTGCTCAAAA	CTCTCTATTC	6600
TCTATTCTAA	TACAGTTTTG	AAAGTTCTGT	CATTTCTGTT	TTATAACAAA	GAAATCTAGT	6660
CATTACTTTT	AGTCTATTTT	ACTAAAATTT	AACAGAAGGG	AACTGGTCAG	AACAGATACA	6720
GAAC TAAAGG	CCATGGCTAG	ACCTGCCAAT	TCTGGGTGTA	GAGCCAGTCC	AACACCTGAA	6780
AAGACTCCTG	CTGCAATCGG	AATTCGACA	ACATTGTAGA	TAAAAGCCCA	GAAAAGATTG	6840
AGTAGAATTC	GATGAAAGGT	TTTCTTACTC	ATATCAAAGG	CACGAACCAC	TCCTAAAAGA	6900

435

TTATTGGTTG TCAACACCAA ATCTGCTGAC TCGATGGCGA TATCTGTTCC AGCTCCCAT	6960
GCAATCCCCA CATCTGCTAC ACTAAGGGCA GGAGCGTCAT TGATACCGTC CCCAACAAAG	7020
GCTACTTTCC CTGACTGTTG CAGTTTATGG ATTTTCATGGG CTTTTTCTTC TGGCAAGACG	7080
CCTGCAATGA CCTCTTCAAT TCCGATTTGA TCTGCAATAG CACGCGCCAC ACCAGCATTG	7140
TCTCCTGTCA GCATGACTGT TCGGAGACCA CGTTTTTTTA GCTGACTGAT GGCTAGCTTA	7200
GCATTTTCTT TAGGAATATC TTGCAAAGCA AGCAAGCCTT TGATTTTCATT GTCAACAGCT	7260
AAGAACACAA CTGTCTTAGC TTCTTTTTCT AGTTCTTCTA GTTTATCTTG ATAAGTATTA	7320
GAAATATCCA TGCCATCCAG CATTTTAGCA TTTCCAAGTA AAACCTGTTT TCCATTGATT	7380
CGCCCTGAAA CACCTTTCCC GTGCAAGGAC TGAAAAATTT CAACAGTTTG AAACCAAGT	7440
CCAGCTTCAC TCGCTCGCTT AACGATAGCC TCAGCCAGTG GGTGTTGAGA AGCATCTTCC	7500
AAGGAGGCTG CCAACCCAAA CACTTCTACT TCGTCGCCGA TGACATCTGT TACCACAGGT	7560
TTCCCTTCCG TCAAAGTCCC GGTCTTATCA AAGACAAGGG TTTGAACTTT CTGGATTTC	7620
TGTAAGACAG TTCCATTTTT GAGGAGAACC CCCATCTTGG CACTACGTCC TGTCCCCACC	7680
ATAAGGGCTG TCGGTGTTGC AAGTCCCAG GCACAAGGAC AGGCGATAAT CAAAACCGCC	7740
ACTCCGTAGA GAAGAGAGGA CACAAAGCTA GCTCCAAGCA CAACCACACT ATCCCTGAGC	7800
AAGACGAACC AAACCCAAAA GGTCAATGATT CCTAAAATGA CAACTACTGG GACAAAAATC	7860
CCTGAAATCT TATCCGTCAA GTCCTGAATC GGCGCACGAC TTGTCTGAGC TTTCTTCACA	7920
AAATCCACAA TCTGAGCCAA AACAGTCTCT GAGCCAACTT TTTCTGCTCT AAAGACAAGC	7980
GTTCACATAT GATTGATGGT TGAGCCAATG ACAGTATCTC CAACTGTCTT GTCCACAGGC	8040
AGACTCTCAC CTGTCACCAT GGATTCGTCA ATACTAGAGA CACCTTCTAC TACGACACCA	8100
TCAACAGCAA TCTTTTCACC GGGACGCACT CGAATCAGGT CGCCTACCTT GACTTGTTCC	8160
AAAGGAACCT GGACATAACT ATCATCACTC AAGACTTCTG CGGTTTTAGC TTGCAAGTCC	8220
AGTAATTTCT CCACAGCTTG GGACGTATTT TTTCTCATTT TTTCTCATAA AACTGCTCCC	8280
AAAAGAACGA AAAAGAGGAT AAATCCAGCA CTTTCGAAGT AAACAGGGAG ACCAGCAAAG	8340
AGAGCAACTA GGCTATAGAA ATAAGCCACT AGAGTTCCCA GCGCAACCAA GGTATCCATG	8400
TTGGCATTGT GCTTTTAACT ACTGGCCCAA GCACTCTGGA TATATGGCTT ACCTGCAACT	8460
AACATAATAG GCGTTGTTGC TAGAAAGGTT CCCCATGCA TGACTTGATG ACTAATGCTA	8520
CCTGTCAACA TCCCAATCAT GAGAATCACA AGAGGCACAG TAAAGATACT AGTAATCCAA	8580
AAACGTTGCA GGAGAGATAG AGATTTTCGA GTCTTCTCAA CGACTGTATA GCTTCCCTTT	8640

436

TGCATCTTCA TGCCACAAGA AAATTCATGT CGCCCTAATT CTTGAGGCGT AAAACGAATG	8700
ACTTTCTCCT CATCTACGCC GATTGGTTCC AAGATACCTT CTTCTTCAAA CAGAATTTCC	8760
TTATAACAGT TTGAAGGAGT AGCACGATGA AAGGTAATCT CAGCTGGAAT TCCCTTTTGA	8820
AGCTGGATAT GGGCTGGATG ATAGCCTTTT TCAGCTCGGA TACGGATTTT TTGAATGCCA	8880
TTTTCTAAGC TTGCTTTCAC AATTTCTGTC ATAGTCTCCA CCTACTCTAC AATCATCTTG	8940
CCGTGCATCA TGTTCATACC ACAAGCAAAG CCAAACTCTC CAGCCTGTTC AGGCGTGATT	9000
TCCACTACAT ACTCTTCCCC CATTTGCCAGG TTCGCATGTA CACCAAAATC TGGAAAAACA	9060
ATTTGATCCA GACATGGTGA AGGATCCTTG CGGTCAAAGA CAATGCGTGC TGGCACTGAT	9120
TTCTTGAGGA CAATCAACTC AGGAGTATAG CCTCCCATGA CTTCCACTCG AATCTCTTGG	9180
TATCCGTTTT TTTGCTGGGC TTTTGTCCA GATTTTTCAG GCTTTTGA AAACCAAAAC	9240
AAGATAAACG CGATAAGGGC AATACAAATA ATGGTTACAA TACTATTTAA CATGACGTCT	9300
CCTTTACATA CAATTACATC TTACTTCTGT TACAGCACTT GATTTCTTCT CTGAAATCAC	9360
AGCTTCCAAG TCTTCCAAGT CAGTCTGAGT AAATTCACAT TCTACAATCA AGTCAGCCAA	9420
CAAAATCCTA ATCCTACGGG AACAAACCTT GTCCTTGATA TCTTGACAA GTAATCCCG	9480
ACTTTGGTCT AGAGTTAAAA GGGCTGAATA AACAAAGGAC TTGCCTTCTT TTTTCCGAGT	9540
CAAACACTCT TTATCAACCA GACGAGCCAA AAGTGTCTGA ACCGTGGACT TGGACCAGTC	9600
AAACCGCTCT GCCAAAACCC TAATCAAATC TGTACTGGTC TGCTCCCCCT GCATCCAAT	9660
AATCTTCATG ACCTGCCATT CTGCATCTGA AATCTGCATT ACCATACCTC CAAAATCTAC	9720
ATTTGTCAAT TACACTCATC AGTATACTCT TAAATCTAC ATTTGTCAAT TATAGAAATA	9780
ATATTTTCTT CGAAAAATAG AATTTTAATC ATTTGAAAAA CGATTTGCAG TCAAATATTA	9840
CTATATAAAC AATAAAAAATA TGCTATACTA AAGAAAAAAG AAAACAACCA CTAGGGGTGC	9900
GTAAAGCTGA GATTAACGAC TGTTAGATCC CTCTGACTCA ATCTAGGTAA TGCTAGCTGA	9960
TGGAAGTGGA AATGATAATG GGGACTAGCA GTCTTCTATT GCCTTTCTAA AACAGACTAG	10020
CTTGTCTTGA AGAATACAAA CTTCACTTGG TTGGGAGGTT TTAGATGACT TATTTACCCG	10080
TTGCTTTGAC CATTCGAGGG ACTGACCTTA GTGGTGGTGC TGGCATTATG GCAGATTTAA	10140
AGTCATCCA AGCGAGAGAT GTCTATGGAA TGGCTGTTGT AACCAGTCTT GTCGCTCAAA	10200
ATACCAGAGG TGTTCACTA ATCGAGCAG TTTCTCTCA AATGTTGAAA GCCCAATTGG	10260
AGAGTGTCTT TTCTGATATT CCACCTCAGG CTGTAAAAAC TGGAATGTTG GCTACTACTG	10320
AAATCATGGA AATCATCCAA CCCTATCTTA AAAAAGTGA TTGTCCCTAT GTCCTTGATC	10380
CTGTTATGGT TGCTACAAGT GGAGATGCCT TGATTGACTC AAATGCTAGA GACTATCTCA	10440

437

AAACAACTT ACTACCTCTA GCAACTATTA TTACGCCAAA TCTTCCTGAA GCAGAAGAGA	10500
TTGTTGGTTT TTCAATCCAT GACCCCGAAG ACATGCAGCG TGCTGGTCGC CTGATTTTAA	10560
AAGAATTTGG TCCTCAGTCT GTGGTTATCA AAGCGGACA TCTCAAAGGT GGTGCTAAAG	10620
ATTTCTCTT TACCAAGAAT GAACAATTTG TCTGGGAAAG CCCACGAATT CAAACCTGTC	10680
ACACCCATGG TACTGGATGT ACCTTTGCTG CAGTGATTAC TGCTGAACTA GCCAAGGGCA	10740
AGAGTCTTTA CCAGGCAGTT GATAAGGCCA AGGCCTTTAT CACAAAAGCT ATTCAAGATG	10800
CCCCTCAACT CGGTCATGGT TCTGGTCCAG TCAACCATAC AACTTTTAAA GATTAAGAAA	10860
AAAACTCTC TAGTTCCAC TTAAAGGAA TTAGAGAGTT TTTATACTCT TCGAAAATCT	10920
CTTCAAACIA CGTCAGCTTC CATCTGCAGC CTCAAACAC TGTTTTGAGC TGACTTCGTC	10980
AGTCTTATCT AAAACCTCAA GGCAGTACTT TGAGCAACCT GCGACTAGCT TTCTAGTTTA	11040
CTCTTTGATT TTCATTGAGT ATTAATTAGG AAAGAATGTT ATGCAACTTT TTTAAAAAGG	11100
CTTGCGTTTT TGCTCAATA TCTTCTGCTT GCATCAAATC ACGTACAACA GCTACACCAG	11160
CTATGCCAGT GCCCATAAGC TGATCAATAT TCTCCGAAGT CAAGCCTCCA ATAGCAACTA	11220
CTGGAATGGC AACCGTTTGG CAAATTGTTT TCAAGGTCGA TATCAGAGTA ATGGGCGCAT	11280
TTTCCTTGGT GGTGGTTGGG AAAATGGCTC CTGTACCCAA GTAATCTGCA CCTGATTCTT	11340
CCGCTTCCAG AGCTCTTTTA ACCGTTTGTAG CGGTGACACC GAGGATTTTT TCAGACCCA	11400
AGACTTTGCG AGCTACCGAA ACTGGTAATT CATCATCTCC GATATGCAGA CCTGCTGCAT	11460
CAACCGCAAG ACAAACATCC AACCGATCAT CGATTATCAA GGTACCTGA TAAGCATCTG	11520
TTATTTCTTT GACTTGTTTT GCCAGTTGAT AATATTGATT GGTGTGAGA TTTTTTCTC	11580
GCAATTGGAC TATGGTAACC CCTGAACGGC AGGCCGTCTC AACTTTTGCA AGAAAGCTTT	11640
CCACGGAATC TTGATAGCGA TTGGTTACCA GATATAGTCT AAGTGCTTCT CTATTCATAA	11700
ACCTCTCCTT TGATGGTATC TAGCCAATTT TCATCTCTTC TTAGGAGCGA AAGCTGATTG	11760
AGTACTTGGT AACGAAATTC TTCCAATCCC ATTCTTGAA CAACTATTTT CTCAGCAGCG	11820
ATATTGAGAT AAGAGACTGC TAAGCAAGAA GCTTCAAAC CAGTCTTTCC TTGGCTGAGA	11880
AAAACAGCTG TTAAGGCTCC AACCAAGTCT CCTGTCCCTG TTATCCAGTC TAATTCAGTA	11940
CAGCCATTTT CCAGTACAGC GACCTGATTT TTCGAAACGA CGAGGTCCTT GGGACCTGTG	12000
ACTAAGAAAG ACATACCAGG ATAGGTCTGA CACCAGTCTT TCAAGACTTG AAGCAAATCC	12060
TCCGTTTCTT GATCTTTAGC ACTCGCATCG ACCCCAACGC CGTGGTGCTT TAATCCAACA	12120
AGACTTCGAA TTTCTGACAT GTTTCCTTTA AGGACCGTAG GTCTATAGTC TAAAGGTCT	12180

438

TTAACTAAGC TCTTACGAAT GGATGAAGTC GPTACGCCAA CCGCATCTAC TACCATCGGG	12240
AGAGAAGATT GGTTCGCATA CGAAGCTGCC ATGCGGATTG CTTTTCCTT CTCAGCTGAC	12300
AAATGCCCCA AATGATGAA GAGAGCCTGA CTTTGCTTAG TAAAATCAAG AACTTCACGG	12360
GAATCATCTG CCATGACAGG TTTGCATCCC AGAGCCAAAA TCCCATTTGC CAGCATCTCA	12420
CAAGAAATCT CATTGGTAAT GCACTGAATG AGGGAAGTAG AGCCTATAGG AAAGGGATT	12480
GTAAATTCTT GCATCAGTCT ATCCTTTCAC TAAAGAAATA TCCCTGCACT TTTTAAAGA	12540
ATTCTGCTT GATTAATAA CGAAAGGCAA TAAAGGAAAT CGCTGTACCA ATCAAGGTTG	12600
CTCCGAAAAA TCGAGGCGTG TAGATAAACC AGCTAAGCTT AGCAGCTGAT CCTGTAAAGA	12660
GTACCATAAC AGGATAGGAA ACAATGGAAC CAATAATACC TGTCCCAAA ATCTCTCCTA	12720
GAGCAGAATA GTGAAATTTT CGACCGTACT TATAAAGAG ACCTGCTAGA AGGGCTCCAA	12780
AAGTCGCTCC TGTGAGAGCT AAAGGCGGAA TCCCTTGAGT CGTCATACGG ATAAAGGCTG	12840
TGACTGTAGC CATAGCCAAG GCATAAACAG GTCCCATCAT GATTCTGCT AGAATATTGA	12900
CTACACTGGA CATCGGTGCC ATTCCCTCAA TCGAAAGAT AGGTGTAAG ACTACATCAA	12960
GGGCAATCAT CATAGATAAA ATGGTTAATT TGTGAACTTG TAATTGGTGC TTTCTCATGC	13020
TTCTATTCTT CTCTTTTTC TAAAGACTGT AAATCGCTCT TCCATGTCTG GTGTTGGTAG	13080
GCCATTTCCC AAAACTTGGC TTCCATATGA AACTGATGT GGAAGGCATC TAGCATTTTT	13140
TGCTTGCTG TCTCGTCACT TTCTCGATAG AGCTGATTGA CCACTGCTCC CTCCTCTCTG	13200
ATCTGTTGCT CTAACATC CGTAATATAA GTTCAATCC ATTGTTGATA GAGAGGATTT	13260
GGTGATGGTT TAAGATTAAG TGATTTGCCT ATATCATGGT ATAACCAAG ACAAGGAAGC	13320
AAGCTTGCAA AAGCGATGGC TAAGTTCGGT TCTGCAAATT GCCTATAAAT ATGAGAAATG	13380
TAATGATAAC AGGTGGAGC GATTGGATGT TGCTCCATTT CCTGGTCGCT GATTTCCAAT	13440
TCCTTGAAAA ATTGTTGGCG AATAAATAAC TCACCCTCCA CTAAACCCTG AGCATTTTGT	13500
TTCAAGAGTC TTTTCATCTC TTGGTTTGAA GTCTTATCAG CCAAAGATG ATAGATTCT	13560
GAGAAAGCCT TCAGATAGTA GGCATCCTGA ATCAGGTAAT AGCGGAAAAAT GGCAGGTTCT	13620
AAATCCCCCT CTTGTAATTG TAAAATAAAG GGATGATGAA AGGAAGCCTG CCAAGCTTTC	13680
TTGGATAATT CCATCGCAAT ATCTGTAAAT TCCATAATAA CTCCTTTATA AAAATAGACT	13740
GGTTTGAGC AATAAAAAGA AAAGCAGGTA GATTAATTTT GTTTTTTTAG GAATATAAAA	13800
AGTCCGATAG CTATTCTTCA ACTGTGCATG TTCGTCATAT CCGTGAGCAG ATAGAGCTCT	13860
CAGGTAAAGA TGGCGCCACC TAAAGACTGT CATCAGAACC TTAAGGTAAA TCAAGGGCGA	13920
CCAAAAATGT AGTTCTTGAC CACGTAATAG GCAAGCTTCT TTGAGGGACT TGATTTCTTG	13980

439

CTGAATGAGA GGAAAAGAAT TGAATACCAC AATCAAGGCA TAGGACCAAG AGCGTGATAG	14040
CCCCTTTGA GCCAAGTACA AGAGAAGCTC TTTTAGTGAA ACAGAGGAAA CAAAGACAAG	14100
GCCGATACAA ACTGTCACAA AGGCCCTCGT TCCAAGCATG ACTGCCTGTG AAGCATCTCC	14160
GTGTAAGTGA ACTGCCCAGT AGTTGGCAAA AGATGGTAAA ATGGCAAGTA TGATCATCCA	14220
AGCTAACATT TTAATCGAC GGTAATAGAG CATAAAGAGA ATACAAAATG CGACTACCGA	14280
AAGAGTCAGA GCAATCGAAG GAATGAAAGA TGTTCCTCAAG GATAAAATCA GCAAGAAGAG	14340
ACTGATAATC GGTGTCTGGG TTGCTACTTT GACCATACTA TCTCACCTCC CCTTGGGTAT	14400
TGCTACTCTG AGATGTAAGT GGTTCGGTAA TGCTCACTTC TTTCACATGC CGAAGACCC	14460
GACTAGTCAT CTCAATCCAA TAATCAACCA CAGAAATCAA AGGGTCTAAA CGATGACTAA	14520
TGAGCAGAAA ACTTCTTCTT TGATTCCTCT CCTCCACAAT CCCTTGCAA AAATAATGGC	14580
AGGCTCTATC ATCCAAACCT GCAAAAGGT CATCTAGCAA GATCACGGAA GCCTTACTGG	14640
TCAAGATGGT CAGGAGCTGA AGAATTTTTT GCTGACCACC ACTTAATTGA TAGGGACTCT	14700
TATCGACTGC CTGCTCCAAA TCAAAATATC GTAAAGCTTG AAAAATCCGC TGATTCTTT	14760
CAGAATCAGG TCCATCTAAT TGAAGCTCCT CTCGCAGACT GACTCGGATA AACTGCTTCT	14820
CAGCTTCCTG AACAACACCA GTCAGATCAC GATACAACT CTTTTCTTT TTCAGGACCG	14880
AACCTTCCA AGTAATGCTC CCCTTATACT TTTGAAATTG AAGAATAGAC CGAAAGAGGG	14940
TTGATTTCCC GACACCATG TCACCCAGGA TACAGGAAAT CCCTTGATAG AATGTGAAAT	15000
CAGCAATTGA AAAGAGGGG CGATTACCAA GCTCACCAGT CACACGGTTC ATATGGAATA	15060
GTTCGGGGCT AGAAGCAACT TCCTTTGAAG CAACCTGTGT CATCTCATAG GAAGGGATT	15120
GAAACACTTC CTTAGTTTT CCGTCTCTTA GCTCCACCAT ATGGTCGATA TAGGCTTTAT	15180
AGTCAGATAA ATCATGGTCG CACAAAATAA CTGCTTTCCC ATCATAGACC AACTCTTTTA	15240
GAATCTCAA TATCTCGATT CTGCTCTTGC GGTCAATGGA AGCGAAGGGC TCATCCAAGA	15300
GATAGACCTT AGGATTCTATG GCAAAGAGGA CAGCCAGCGC TGCTTTTTCG TTTTCCCCAC	15360
CTGATAAGTG ATGGATGAGA CGGTGCAAGA TGTCTTTGCA ACGACATTGC TGGACAACCT	15420
CTGCTATTTT AGAATCAATT TCCTGAAGGT GATAGCCGAT ATTTTCCATG GTAAAAACCA	15480
ACTCTCAAA CAAGCTCTCC ATGCTAAATT GATGATTAGG ATTTTGCAAG AGAATACCAA	15540
CCGTCTGGAC ACGTTCGACG ATAGAAAGCT GACTGACCTC GCTCCCATCT ATCAGGACTT	15600
GACCGCTATA GGGAAGAGAA CTAAGTTGGG CAATCATTTG AAAGAGGCTG GATTTTCCAG	15660
ACCCACTACT CCCAATAAC AAGGTAAAGG CTTGCGCATG AAAAGTAAAA TCAAACGGCT	15720

440

CAGAGAAGAT	TGGGGACTGA	ATCGCTCGTA	GTTCAGACC	CATCTATGCT	TTTCCTCCAG	15780
TTGCAAACTG	ATGATAGAGT	TTGACAATGG	CACGAACCAA	GATGGTACAG	AAGAAATAAA	15840
CAGAAATAAA	ACGTACCACA	AGCAAGGAAA	GGACAAACGG	AAGGGAAAAG	GCGTAGTAAC	15900
CTAACTTAAT	GTATTTCATAG	ACAAAGCTAA	CAAGCGTAAT	CCCAATACTA	TTAGCAGTTA	15960
GAGAGAGCCA	ACTTTCATAG	CGATTCTTAG	TTACGATAAA	ACCAAATTCA	CTTCCCAAAC	16020
CTTGAACAAA	GCCAGACAAA	AGAGCTCCTA	GACCAAATTG	GCTACCATAA	AGGACTTCAG	16080
CAAGCGCAGC	TAGCACTTCT	CCAATCGTTG	CACTTCCGAC	TCTCGGAACA	AAGATGGCAG	16140
CAATGGGCGC	AGCCATACAC	CAGAGACCGA	AGAGGATTTC	ATTGGCAAAG	GCCTGCAAAC	16200
CAAGAGGTGT	TAAGAGTAGA	CTGAGAATAT	TATACACATA	TCCTGAACCA	ACGAAAACCC	16260
CACCAAAAAA	GATAGACAAG	AAAGCAAGCA	AGATAACATC	TTTAACTGTC	CATTTTTCAT	16320
ACATAAAAAA	CTCCTTTTTT	TAAAGAAAAG	TGAGGCACTC	AAGAAGACCG	ACCTAAATAC	16380
TTTGTATAGC	AGACTGAATT	TAGAACAGTA	CACAAGAACA	CTAAAATATT	TCTAGAAATT	16440
AATTTGAATT	TTCTAATTGA	TTTGTTCGCA	TCTTATTTCA	ATCTACTATA	TCATCTTCAT	16500
CCAGTTTCGT	AAAAGAAAAA	ACTCTAATTA	CAGATACAAA	TTAGAGTTCA	GCTTACAAGA	16560
TTAGACAGTT	CTTTTCGACA	TACGAAAAAA	ACATTTTACA	TTTCCCTTCG	CCAGTCTTAA	16620
CTGTATCAGG	TTCAATGGGT	ATCATCTCAG	CCTAAAGCAC	CCCAAATGTC	TTTATTATTT	16680
AATTATGTGA	TTATTATAAC	ACACATTTTA	TACTAGTTCA	AGAAATGAA	CTGGAAATAC	16740
AGCCTTGAC	TCACAAGAC	AGCAGATCTT	TCTTTTGCAA	AAAACAAATG	ACCTGTTTGA	16800
TGAATTAGCC	ATTCAAGCTG	AATCTGGACA	TAGCTTTTTA	AAAAGGAAA	ATCCTACTTA	16860
CTTAGAATCC	AAGGATAGAT	ATCTATTGTT	CACTCATTTC	CCGAACAGTT	TTTTCTATAT	16920
TTTTTGCATA	CGATATTGCC	GAAATGATTG	AAACGCCATC	CATATTGGTC	TTTATAATGT	16980
CTTTAATATG	TTTCGTCTGT	ATCCACCAA	TTGCAACTAA	AGGCATTTGT	GGCAATAGTT	17040
TTCTCATCAA	TTCAAGACCT	TCATAACCTA	TAGTACCACC	AGCATCATCC	TTTGACTGGG	17100
TACCAAATAC	AGGCCCAACA	CCTACATAAT	CTACATATTC	AACTTTTGAT	TGTTGAAATT	17160
CTTCTTCGTT	TCTTATAGAA	AGACCAATTA	TTTTATCTGG	CATCAATTTT	CTAATTTTCAT	17220
CAACACCAAT	ATCATCTTGA	CCTACATGTA	CGCCATCGGC	GTCAATTTCC	ATTGCTAAAT	17280
CTATATCGTC	ATTAACGATA	AATGGAACAT	TGTATTTTTT	ACAAAGTTCT	TTAATTTGGA	17340
TAGCTAGCTC	AAGTTTTTCT	AAGCCTTCTA	AAGCACCTTC	ACCTTTTCT	CGAAATTGAA	17400
ATAAGGTTAT	ACCACCTTTT	AAGGCTTCCT	CAACGACTGT	ATATAGATTT	TTTCCTTGGC	17460
AAGTAGTCGT	TCCACAAATA	AAATATAGTT	TTAGTAATTC	TTTATGAAAC	ATCTTACTTC	17520

441

ACTCTTTTGA ATTCTTTTAC ATCTTCATCT GTAATCTCGT ATAAGGCATT TATAAATTCA	17580
ACTTTAAATG TCCCAGGAAG ATGTCCATTT GGACGTTTTT CTGCTATTTC TCCAGCGATA	17640
TTGTAAACCA ACACTGCTGT TTTTAATGAT TTCAATTCTT GACCTTTTTT TAGTCCGATA	17700
AAGCTTGCTA CTACAGCTCC TAATAAGCAT CCTGTCCCAA TGACTTTTCGG CATCATAGCA	17760
CTACCATTAT GAATCATTAC CACTTCTCCA TTAACAGCAA TGGCATCCAC TTCACCTGTT	17820
ACTACTATTG GAATATTGAA CTTCTCATTT GCTGCTAGAG CAATTTCGTC AATATTATCT	17880
ACGCCCCGAC TATCTACTCC TTTAGATGCC ACATCTATTTC CTACTAAAGA GGCAATCTCG	17940
CCAGCATTTT CTCTAATCGC TGCTAGTTTA TAATTGTTGA TTAGATCATC TGCTACTTTT	18000
TTTCTATATT CTCCTGCTCC ACAGGCTACA GGATCTAAAA CTGCTGGGAC ATTATATTTT	18060
TCTGCAATTT TCAGAGCAGC TTGGTATAAT TTCCAATTTT CATCTGTCAA TGTTCCTATG	18120
TTTATTAATA AACCACCAGC ATACTTTAAC AAATCCTCTA AATCTGCTGG AAATCCTC	18180
ATGGCTGGTG AGGCGCCAG TGCTACTAAT CCATTTGCTG TGAAATTTTT TACTACATCA	18240
TTGGTTATAC AAATGACCAA TGGTGCTTTT TCTTTTAATA ATTTTAAACT TGTCAATTTG	18300
AAATCCTTCC TTTTCACTTT ATACGATCTA CTAATTTTGA TTTATCTTTA GTTGAGAATT	18360
TTTTTCATTT ACATTGAATG ATTTATACTC AATGAAAATC AAAGAGCAAA CTAGGAGGCT	18420
AACCGCAGGT TGCTCAAAAC ACTGTTTTGA GGTGTGGAT AGAACTGACG TGGTTTGAAG	18480
AGATTTTCGA AGAGTCTTAC CTCATCAAAT TTGTAAATAT CATGAGCCTT CTCTAGACAT	18540
CGTAACCAAT ATCAAAAAA GCTAATTCTA AAGCGACTGC TTGATTCCAG CGTTGCTGAA	18600
GTTCTGTCAA ATCTTCTCGA TTTTACCAG CACGATTGAG TTCGTCAACC AGAAATTGAA	18660
CCCACTCTGC AAAGAAAGGA CCTCTGTGGA GATTGATCCA TTCCGAATGA ATATAGACTT	18720
CAGGTAAAGC CAAATCTTTA GAACCCAGT CTAAATAGAG ACCTTCTGCA ATGACCAGCA	18780
TGACCAAAAG ATGGGCATAG TCTGATGAAG CCACCGCCGA ATACATTAGA TCCTGAAAGG	18840
CTTTTGTTAC AGGGTGCAA GTCACCTCTA GATAGTCATT CTCTGCTACT TTTAACTCTT	18900
TAAAAGCCTT TTGGAAATAA CCATCTTCAT CTGCTTCAAG AAAGCCTAGT TGCTTGGCAA	18960
AACGAAGCTT GGATTCAAGT TTATCTGCGT GACTACGCAG GCACCCAGCA TGGATAAGAA	19020
GGCATCAAAG AAGTGATAAT CTTGAATCAG ATAGTCCTTT AAGACCTTAT TCTCAATTGT	19080
CCCCGAAAA AGTTCCTTAA CAAAACGATG ATTGATTGCA GCCTGCCAAT CCTTCTGACT	19140
GCTTTTAAAT AATTCTCCAA CAGTCAAACC TGGCTGAAAT GCATAGTCTT GTGTTTCCAT	19200
ATTACTTCTT CCTCTCTTTA CTTGTTAGTA ATTAATAAAA CACCAAGAAA TATCAAGCAA	19260

442

AATCGTAATT CCACTTGATC CTTTAAAGC ACATCGAGAG CATTTGCAGA GAGCTAACTA	19320
AACAAGCCTA TCCAGTTTAT ATAAACAAAA AACTCCAATT ACAATCAAGA ATTAGAGTTG	19380
ACTTACAAGA TTAGACCGTT CATTTACCA TACGAAAAA CTGTTACAT TTCCTTCGC	19440
CAGTCTTAAC TGTATCAGGT TCAATGGGT TATCTCAGC CTAAAGCACC CCAAATGTCT	19500
CTATTATTTA ACTACTGAAC CAGTATAGCA AAAAATGAAA GCCCTAGCA GATATTTGAC	19560
CGAAAAATAT CTTTATATAT AATATATTGA AACTAGAATA GTACACCTCT ACTTATAAAA	19620
CATTGTTAGA AATCGATTG ACTGTCCTGA TTGATTTGTC CTATTCTTAT TTCATTTTAC	19680
TATAGTTTTC GATAGCAATT TATTCTTCCA ATACACGAAG AAAACCTCC ACATTCAGTG	19740
GAGGCAATCT GTTTTATCAA TACAATTTA AGTCACGAGG GTCAACTGGG AAGGTTGGGT	19800
TGTATGGATT GTGACGGAGC TTGAAGTGT TGACATCTTC AATGGTCTGA GTTCCAGACA	19860
ATTGCATAAC TGTCTTCAAT TCCGCATTCA AGTGTTCAAA GACTTGACGC ACACCGACAC	19920
TACCACCGAG AGCCAAGCCA TAGATGACAG GCGCTCCAAT AGCAACCAAG TCTGCTCCTG	19980
ATGCCAAGGC TTAAAGACG TGTGACCAC GACGAACACC AGAGTCAAAG ACAATCGGCA	20040
CACGTCTATC AACTGCTTCT GCCACTTCTT GAAGCGAGTC AAAGGCAGCT GGTCCACCGT	20100
CGATTTGACG ACCACCGTGG TTGGTTACCC AGATACCAGA AGCTCCTGCA GCAAGCGAAC	20160
GTTCAACGTC CTCACGGCAT TGTGGTCCCT TGACATACAC AGGAAGACCA GAGTATTCAG	20220
CGATAAATTC TACATCGCGT GGAGACAAGC GTTGTTTAGC TGATTTGTAA ACAAGTCCA	20280
TTGATTTACC AGCACCTTCT GGCAGGTATT CTTCAACAAT CGGCATGCCA ACTGGGAAGA	20340
CAAAACCATC ACGCTTATCC ACTTCACGAT TCCCCCTAC AGTAGCATCT GCCGTCAAGA	20400
CAATCGCTTT ATAACCTTCA GCCTTCACAC GGTCCATGAT GTGGCGGTTG ATACCGTCAT	20460
CCTTACTAAA GTAAAATTGA AACCAATGAG GTGTCCCTTG GAGGGCTTCA GAAATCTCTG	20520
GAAGGTCAAC AGTAGAGTAA GAACTGGTTG TATAAAGAGA ACCAAACTCA TGCACACCAC	20580
GCGCAGTCGC CACTTCCCC TGTTCATTG CCAATTTATG AGCCGCAACA GGTGCCATAA	20640
TGATTTGAGA AGATAGTTTT TCACCTGCAA ATTCAATCTC TGTACTTGA TTTTCTACAT	20700
TGCAAAGTGT ATGAGGAACG ATGAGCTTGT GGTAAAGGC ACGGATATTC TCTCTTAAAG	20760
TGAAAGTATC TTCCGCCCCA CTAGCGATAT AGCCAAATGC TGCTTTAGGA ATAACTTGT	20820
GCGCCATTGG CTCCAAATCA TAGGTATTGA TGAATCTAC ATGACCTTCT GCATTGCTTG	20880
TTTTGTATGA CATAAAATGT CCTCCTTAAT AAGTAAGCGT TTACTTTGTG TATTACAAAA	20940
ATATCTTAAC TCTTTTTCAA AACTTTTAAA ATATTTTGTG TGGAAATTC AGAAATTTTA	21000
TGTCTATGAT AAAATCCTT ATAACGGCAA TAAAAATAG ATATTATCCA AAGAAGATTT	21060

443

TAAGTGCTAC AATAACTGTA TTATTTCTAG ATGGGAGGTT CTATTTTGG ATTGATCCAT	21120
TGTTGAACAA TATCTACCAC TATATCAAAA GGCATTCTTT CTGACCTTGC ATATTGCAGT	21180
TTGGGGAATT TTGGGATCCT TTCTGCTCGG TTAAATCGTT AGTATCATCC GACATTATCG	21240
AATCCTTGTT TTGGCGCAAG TAGCGACAGC CTACATTGAA TTGTCACGTA ATACGCCCTT	21300
TTTGATTCAA CTCTTCTTTC TCTACTTCGG TCTTCCCCGA ATCGGGATTG TCCTATCTTC	21360
AGAAGTCTGT GCAACGCTTG GGCTTGTCTT TTTAGGAGGC TCCTATATGG CAGAACTTTT	21420
CCGAAGTGGG CTGGAAGCCA TCAGTCAAAC CCAGCAGGAG ATTGGCCTCG CTATTGGTCT	21480
GACACCTCTA CAGGTCTTTT ACTATGTGGT TCTTCCGCAA GCAACAGCGG TGGCACTCCC	21540
CTCCTTTAGT GCCAATGTCA TTTTCTTAT CAAGGAAACC TCTGTTTCT CAGCAGTGGC	21600
TTTGGCCGAC CTCATGTACG TCGCCAAGGA TTTGATTGGT CTCTACTATG AGACAGACAT	21660
TGCGCTAGCT ATGTTGGTAG TTGCTTATCT AATCATGCTG CTACCCATCT CACTGGTCTT	21720
TAGCTGGATA GAAAGGAGGC TCCGCCATGC AGGATTCGGG AATCCAAGTA CTCTTTCAAG	21780
GAAATAATCT CCTGAGAATC TTACAGGGAT TGGGCGTTAC GATTGGGATA TCCATCCTGT	21840
CTGTCTCTT ATCCATGATG TTCAGAACAG TCATGGGAAT CATCATGACC TCCCATTTCTA	21900
GAATCATACG ATTTTAAACA CGATTGTATC TGAATTTAT CCGTATCATG CCCCAGCTGG	21960
TGCTACTCTT CATCGTTTAC TTTGGCTTGG CTCGAACTT TAATATCAAT ATCTCAGGTG	22020
AGACTTCAGC TATTATCGTT TTTACCCTCT GGGGAACAGC TGAAATGGGA GACTTGGTAC	22080
GTGGAGCTAT CACTTCTCTC CCTAAACATC AGTTTGAAAG TGGACAGGCA CTCGGCTTGA	22140
CTAATGTTCA ACTTTACTAC CACATCATCA TCCCACAAGT CTTAAGAAGA CTGCTACCGC	22200
AGGCTATCAA TCTTGTCACG CGGATGATTA AAACCACTTC ATTAGTTGTT TTGATTGGGG	22260
TTGTGGAAGT GACCAAAGTT GGACAACAA TCATCGATAG CAATCGCCTG ACCATCCCAA	22320
CTGCTTCATT TTGGATTTAT GGAACCATTC TAATCTTATA TTTCGCAGTT TGCTACCCTA	22380
TTTCCAAACT ATCCACTCAC TTAGAAAAAC ATTGGAGAAA CTAAATGTCT GAAACTATCT	22440
TAGAAATCAA GGAACAAAA AAATCCTTCG GAGACAATCC CATCCTCCAA GGACTTTCTC	22500
TAGAAATCAA AAAAGGGGAA GTTGTGTGCA TCCTAGGGCC ATCTGGTTGT GGGAAAAGTA	22560
CCCTCCTTCG TTGCCTCAAC GGCTTAGAAA GTATTCAAGG TGGAGATATT CTCTGGATG	22620
GTCAGTCTAT CGTTGAAAAA AAAAAAGATT TTCACCTAGT TCGCCAAAAG ATTGGCATGG	22680
TCTTTCAAAG TTATGAACTC TTTCCCCATC TGGATGTCTT ACAAACCTC ATCCTAGGCC	22740
CTATCAAAGC TCAAGGAAGG GACAAGAAAG AAGTAACGGA AGAAGCTTTG CAATTACTAG	22800

444

AGCGTGTCCG	TTTGCTGGAT	AAACAACATA	GCTTTGCCCG	TCAATTATCT	GGTGGACAGA	22860
AGCAACGTGT	TGCAATTGTC	CGTGCCCTCC	TAATGCATCC	AGAAATCATC	CTTTTGTACG	22920
AGGTGACTGC	TTGCTGGAT	CCAGAAATGG	TGCGTGAGGT	GCTGGAAC TT	ATCAATGATT	22980
TGGCCCAAGA	AGGCCGTACC	ATGATTTTAG	TAACCCACGA	AATGCAGTTT	GCCCAAGCCA	23040
TTACTGACCG	GATTATCTTC	CTCGACCAAG	GGAAAATCGC	TGAAGAAGGA	ACAGCTCAAG	23100
CCTTCTTTAC	CAATCCGCAA	ACCAAACGAG	CCCAGGAATT	TTTAAACGTC	TTTGACTTTA	23160
GCCCAATTCGG	CTCATATCTA	TAAAGGAGAT	TCTTATGAAA	CTATTCAAAC	CACTCTTAAC	23220
TGTTTTAGCA	CTTGCCTTTG	CCCTTATCTT	TATCACTGCT	TGTAGCTCAG	GTGGAAACGC	23280
TGGTTCATCC	TCTGGAAAAA	CAACTGCCAA	AGCTCGCACT	ATCGATGAAA	TCAAAAAAAG	23340
CGGTGAAC TG	CGAATCGCCG	TGTTTGGAGA	TAAAAAACCG	TTTGGCTACG	TTGACAATGA	23400
TGGTTCCTTAC	CAAGGCTACG	CTACGATATT	GAAGTAGGGA	ACCAACTAGC	TCAAGACCTT	23460
GGTGTCAAGG	TTAAATACAT	TTCAAGTCGAT	GCTGCCAACC	GTGCGGAATA	CTTGATTTCA	23520
AACAAGGTAG	ATATTACTCT	TGCTAACTTT	ACAGTAACTG	ACGAACGTAA	GAAACAAGTT	23580
GATTTTGCCC	TTCCATATAT	GAAAGTTTCT	CTGGGTGTCG	TATCACCTAA	GACTGGTCTC	23640
ATTACAGACG	TCAAACAAC T	TGAAGGTAAA	ACCTTAATTG	TCACAAAAGG	AACGACTGCT	23700
GAGACTTATT	TTGAAAAGAA	TCATCCAGAA	ATCAAAC TCC	AAAAATACGA	CCAATACAGT	23760
GACTCTTACC	AAGCTCTTCT	TGACGGACGT	GGAGATGCCT	TTTCAACTGA	CAATACGGAA	23820
GTTCTAGCTT	GGGCGCTTGA	AAATAAAGGA	TTTGAAGTAG	GAATTACTTC	CCTCGGTGAT	23880
CCCGATACCA	TTGCGGCAGC	AGTTCAAAAA	GGCAACCAAG	AATTGCTAGA	CTTCATCAAT	23940
AAAGATATTG	AAAAATTAGG	CAAGGAAAAC	TTCTTCCACA	AGGCCTATGA	AAAGACACTT	24000
CACCCAACCT	ACGGTGACGC	TGCTAAAGCA	GATGACCTGG	TTGTTGAAGG	TGGAAAAGTT	24060
GATTAGTCAT	TAACTCTTAA	AAGGAACTGG	ATTTTAAGCT	CCAATCCCTT	TTTAAGATTT	24120
TACCTATAAC	ATCCTGAGTC	TATCTAAGAT	GTTCAATCTG	AACACAGTGT	ACATACTTTA	24180
TCTTCTATTG	CATATACTTT	ATCACATAAG	ATACGAATAT	CCTCTTCACT	ATGACTAGCA	24240
ATCAAAATTG	TTGTCCCTTT	TTCACTAGAG	AGCTTTCTAA	ACAATGTTCT	CATATTTTCT	24300
ACACTTGATT	TATCCAAGGC	ATTCATAGGT	TCATCTAGTA	AAAGAATAGA	GGGATCTCTC	24360
ATAATTGCTT	GAGCAATCCC	TAGCTTTTTC	CTCATACCTA	GCGAATAAGT	TTTAACTTTC	24420
TGGTCTTTTT	GCTCATATAG	ACCAACTATT	TTCAGTGTAT	CATTGATTTT	CTGATTACCA	24480
ACTACTCCTC	GTATGCTTGC	CAAATATTGT	AAATTCTTAA	AGCCACTATA	ATAATTTATA	24540
AAACCAGGTT	CTTCAATCAA	AGCTCCCAAA	TTAGCTGGAA	TTTTTCTCTC	AGGAACAATA	24600

445

TTTTCCCAT TGATTAACAC TTCTCCATAA GACGGACTAT ATAAACCAGC TATTAATTTA	24660
AACAATACAC TTTTCCCTGA GCCATTGCGA CCAGTAATTC CTATAATTTT CCCCTGTTTA	24720
CAACTAAAGT TAAGGTTTTG AAAAACACAT GTCTTTTTTA ATTTCAACTC AATATTTTTT	24780
AATGTAATTA TTTTCATTCAT TCTATAAACC TCCTCTTTTG ACGAGTGAAA TAGAAAATGC	24840
TTTGAAAAG AAAGACTAAA AATAGCAACT GAAGAAATAA ATCTCGTCCT ATATCTCCAT	24900
TCCCTCGATT CAAAATATAA AATAGATAAT TAGTTCGATT TCCTACAAAT AGACCACCAA	24960
ACACAATCAT GAGTAAAAG AAACAAACGC AAGCAAAGTT CG	25002

(2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11443 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

CAGGTACGGT GAGGCGCAAC TAAATATAA TTTTCATCTT GATTAGGAAT TTTATCAGTA	60
TTATGATAGT GAGCATTGCC ATTGATGGAC CATAAGAGCA ATACAACTAA TCCACGCAAA	120
TAAGTATAAA ACATGCGATC TCCTTCGATT GTTTCTTGT TATTATTATA CCTTATCAAA	180
GGAGGGCTGG CAAACTTTTC CCTTGACTAG ATACATATTT AGGATGAAAT TAGAATTCTG	240
TTAAAAAAA TGATATAATA GAATTTATGG ATAAAAATAA GATTATGGGA TTAACCCAAA	300
GAGAAGTCAA GGAAAGACAG GCTGAGGGT TGGTCAATGA CTTTACCGCA TCAGCCAGTA	360
CCAGCACTTG GCAAATCGTT AAACGAAATG TCTTTACCCT TTTTAACGCT TTGAACTTTG	420
CCATTGCTTT GGCTCTTGCC TTTGTGCAGG CTTGGAGCAA TCTGGTCTTC TTTGCTGTTA	480
TCTGCTTTAA CGCTTTTTCT GGGATTGTGA CCGAGCTACG AGCCAAACAC ATGGTGGACA	540
AGCTCAATCT CATGACCAAG GAAAAGGTCA AAACCATCCG TGATGGTCAG GAAGTTGCTC	600
TTAATCCTGA AGAATTAGTG CTAGGAGATG TCATTGCTTT GTCTGCAGGA GAGCAGATTC	660
CTAGTGATGC CTTGGTTTTG GAAGGCTTTG CGGAAGTCAA TGAAGCCATG TTAACGGGAG	720
AAAGTGATTT GGTGCAAAAG GAAGTTGACG GCTTACTTTT GTCAGGAAGT TTCCTAGCCA	780
GTGGGTCACT TTTATCTCAA GTTCACCATG TCGGTGCAGA CAACTATGCT GCCAAACTCA	840
TGCTTGAGGC TAAGACCGTT AAACCCATCA ACTCCCGTAT CATGAAATCG CTGGACAAGT	900
TGGCTGTTTT TACTGGGAAG ATTATCATTC CCTTTGGTCT GGCTCTCTTG CTGGAAGCCT	960

446

TGCTTTTAAA AGGCCTGCCT CTCAAGTCAT CCGTTGTAAA CTCGTCGACA GCTCTTTTGG	1020
GAATGTTGCC TAAGGGAATT GCCCTTTTGA CCATTACTTC GCTCTTGA CTGAGTGATTA	1080
AGTTGGGCTT GAAAAAGGTC TTGGTGCAGG AGATGTACTC TGTGAGACC TTGGCGCGCG	1140
TGGATATGCT CTGTCTGGAC AAGACGGGTA CCATCACCCA AGGAAAGATG CAGGTGGAGG	1200
CTGTCTTCC GTTGACGGAA ACGTATGGTG AAGAGGCTAT TGCCAGCATC TTGACTAGCT	1260
ACATGGCCCA TAGTGAGGAT AAGAATCCAA CTGCCCAAGC CATTGCGCAG CGTTTGTGG	1320
GAGATGTTGC TTATCCTATG ATTTCCAATC TTCCCTTCTC GAGCGACCGC AAGTGGGGGG	1380
CTATGGAGTT AGAAGGCTTG GGGACAGTTT TCTTAGGGG ACCTGAGATG TTGCTTGATT	1440
CTGAAGTCCC AGAAGCTAGG GAGGCCTTG AGAGAGGATC ACGTGTCTTG GTCTTAGCTC	1500
TCAGTCAGGA GAAATTAGAC CATCACAAAC CACAGAAACC ATCTGATATT CAGGCTCTAG	1560
CCTTGCTGGA AATCTTGAC CCCATTGAG AGGGAGCAGC AGAGACGCTG GACTATCTCC	1620
GTTCTCAGGA GGTGGGACTC AAGATTATCT CTGGTGACAA TCCAGTTACG GTGTCCAGCA	1680
TTGCCAGAA GGCTGGTTTT GCGGACTATC ACAGCTATGT AGATTGCTCA AAAATCACCG	1740
ATGAGGAATT GATGGCCATG GCGGAGGAGA CAGCTATTTT CGGACGTGTT TCCCCATC	1800
AAAAGAACT CATCATCCAA ACGTTGAAAA AAGCGGGACA TACAACGGCT ATGACAGGGG	1860
ACGGGGTTAA TGATATCTTG GCCCTTCGTG AGGCGGATTG TTCTATCGTG ATGGCGGAGG	1920
GGGATCCAGC AACCCGTCAG ATTGCCAATC TGGTTCTCTT GAACTCAGAC TTTAATGATG	1980
TTCTCTGAGT TCTCTTCGAG GGTCGTCGCG TGGTCAATAA CATGCCCCAC ATCGCCCCGA	2040
TTTTCTTGAT AAAGACCATC TATTCCTTCC TGTTAGCAGT CATCTGTATT GCCAGTGCTT	2100
TACTAGGTCG GTCAGAGTGG ATTTTGATTT TCCCTTCAT TCCGATCCAG ATTACCATGA	2160
TTGACCAGTT TGTGGAAGGT TTCCCACCAT TCGTTCTGAC TTTTGAGCGA AATATCAAAC	2220
CTGTTGAGCA GAATTTCTC AGAAAATCCA TGCTTCGTGC CCTACCAAGC GCTCTCATGG	2280
TCGTCTTCAG CGTCCTGTTT GTGAAAATGT TTGGCGCGAG TCAAGGTTGG TCTGAGTTAG	2340
AAATCTCAAC TCTACTCTAT TATCTCTTGG GGTCAATTGG TTTCTTATCC GTATTTAGAG	2400
CCTGCATGCC ATTTACCCTA TGGCGTGTCC TCTTGATTGT TTGGTCAGTA GGAGGTTTCC	2460
TAGCCACAGC TCTCTTCCCA AGAATTCAAA AACTGCTTGA AATTTCACC TTAACAGAAC	2520
AAACGTTGCC TGTATTATGGT GTCATGATGT TGGTCTTTAC CGTGATTTTC ATCCTGACCA	2580
GTCGTTACCA AGCGAAAAA TAAATCAAAA CCACCAGTGT GAACTGGTGG TTTGTTCTGC	2640
GGCTATAAGC CGCTTCTACC GGCCAGGGCC AAAGGCCAC CGAAATAGCT TCCTCGCGCA	2700
CCACTTTCCC GAGCAGGTGC TAAAGCACCT TAGTTACTTC CTCTTATTTA TTTGCCAGT	2760

447

AAACGGATCT ACTGACTCGA ATAACGTGAG CTGGTCTGCT ACTCTGTCTT CTTGTAATTG	2820
ATTCTGAATA TATTCAGCTA TCACTTTCTG ATTACGGCCT ACCGTATCTA CATAATAGCC	2880
TCTACACCAA AACTTGCGAT TGCCATATTT GTATTTTAAA TTCGCATGCT TATCAAAAAT	2940
CATCAAACCTG CTCTTGCCCT TTAAATAGCC CATAAAGGAC GAAACACTAA GTTTCGGAGG	3000
AATACTGATA AGCATGTGAA TATGGTCTGA ACAAGCATTG GCTTCATGGA TTATTACACC	3060
CTTACGCTCA CATAAGTCAC GTATGATTCT TCCGATACTA GCTTTGTATC TGCCATAAAT	3120
GATTTGACGA CGATATTTGG GTGCAAAAAC AATATGATAT TTACAATTCC ATGTGGTATG	3180
TGATAAACTT TGATTATCCT CTCTCATGAG GTACCTCCTG TATGATA'GT TGTAGTGGCG	3240
GAGAAACCAC TTCTATCTTA TCATTTTAGG AGGTTCTTTT TGTACCACG CTAAGGCTC	3300
TATGGAA'CCA CTAGCATAGC TAGTGGT'TTT CGGGAGACAA CAAGAAAGAC TGCAATCTGT	3360
GGATTGCAGT TTTTATACG ATGGATCTAT CGTAGATCTG ATGTGCAAGG CCTACGTGCC	3420
GATCATCTAT CGGTGAACCC AAGAGCGACC CTCAAGCCTG CTTGGATTGA GGTAAATAGAT	3480
TCAAATATCT GTAGTTAGAC TATTTGAAGT TTGATGTAAG AAAGAGAAAG CGACAGATTG	3540
AAGTAATTTT AACTCTCTTC TATTGCTAGA ACAAATGGTC GGATAGGTTG GTAGTTTGAA	3600
AATGAAGATG CTATCTATTG TTAAATGGAA CATAGTGTTA TTTATTAGAA AATCGTTTGG	3660
TTTATTCTTT ATCAAATACG AAAAGCAACT TAAATATTTT AACTAAAATA GATGTTATGA	3720
AGAAAAGGTA AAATGATTTT GGCATAGTGA GGTCTCTTTC TATTTGATAT CATATTTT'G	3780
ATAAAAACAA AAATGTCCAT TGCAAAGGAC AAAATGCGAA GTATATTATT TTTTGAAAGC	3840
GATATAATGG ATTCATAAAG GAGGTGTATC GTGTCTAGAA AACAAGAACA AATGGAAACG	3900
TTGTTGCTCC TTTTGCGAGA TAGTAAGGAT TATATATCTG CTAAAGTATT GGGAGAAAAA	3960
TTAAATTGCT CTGATAAAAC GGT'TTATCGC CTTGTCAAGG GAATCAACAA AGATTGTCCG	4020
GTAGAAGCAT TCATTTTATC TGAAAAAGGC AGAGGTTTCA AATTAAATCC AAGAAGTTCC	4080
CTCGTGGACG TTGATGGGAA TTTTACAGAG GCTTTTGATC CTGAAGTAAG GCGTGAAAAA	4140
TTACTAGAAC GTCTCTTGTG GACTGCTCCT AAGCCACATT CTATTTATGA TTTAGGAGAG	4200
GAATTCTACG TAAGCGAGTC AGTAGTACTA AAAGATCGTC AGATATTACA AGAGAGTCTA	4260
GCAATTTATG GGTTAGATTT AAAAAAGAGA CAACGAAAGC TTTTATATGA TGGGGATGAG	4320
GCTCAAATTC GTTCAGCCAT TCTAAATCTA CTGCCAATGT TTAATCAGTT GGATTTAGAG	4380
CAAATTACAC AGAATAAGGT TCAGCCTCTT GACGGAGAAC TTGCTCACTT TTGTTTGGGA	4440
TTACTGATTA CACTTGAGAG AGAATTGGGG GTAAACATTC CCTATCCATA TAATATAAAT	4500

448

ATTTTCTCTC ACCTGTATAT TTTTATCAGT AGGAATCGTC GTAGTACTAG TATTCATGTT	4560
GTAGCACCTT CAAAACCTAC TATTGTTGAT GAGAAAATTT ACAGTGTCTG TCAAAAAATT	4620
ATTCAAGAAA TTGAACAATA TTTTAGGATG AAGGTTGATG CAGTTGAGAT TGA CTATCTTT	4680
TATCAATACG TTGTATCTTC GAGATTGCAA AAACCATTTT CTCCCGGAA GCTTCCTTTT	4740
TCTCAGCGAG TTTTAGATGT CACTCATTAC TATTTTAGCC GTATGTGTAT GGACAATAGA	4800
GAGATTGAAA CGACAGATCC TGACTTTGTT GACTTGGCGA GTCATATCAG TCCCTTACTG	4860
AGGAGATTAG ATAATAGAGT ACAGATTAAG AATAGTCTTT TATCACAAAT TCTTTTAACC	4920
TATCCTAATC TGGTTAAAGA GTTAACAAC TTTTCTAAAG AAGTGAGTCT AGTATTTGGT	4980
TTTGCTTCCT TGAGTCTGGA CGAGATTGGT TTTCTAGTCT TATATTTTGC ACGGTTTCAA	5040
GAAAAGCGAG CACGTCCTCT AAAAACAGTA GTGATGTGTA CATCAGGTGT CGGAACCTCA	5100
GAGCTTTTAC GAGCACGATT AGAAAAGCAA TTTTCTGAAT TGGATATTAT TGATGTAGTT	5160
GCTTATCATC AATTAGATGA GCTGATAAAT CTATATCCAG ATTTAGATTT CATTGTGACC	5220
ACGGTAGCTT TGCAGGAACC AGCAAGTGTC CCGTTTGTCC TAGTTAGTGT TTTTCTAACC	5280
GAGGGTGATA AACACGTCT TCAAGCAAAA ATTCAGGAGA TAACTATGA ATAATCTTTC	5340
GCTTGTCCCT ATGGATATAT CTGTTCAAAA TCGTCAAGAA GCCTACAAAG AATTAGCAAA	5400
TCAAATCAGC CTCTTTGTTT CTGAAGATAC AGAAAAAATA GAAGAGCTTC TATATTACCG	5460
TGAGAGACAG GGAAGTATAG AGGTTGCTAA AGGTGTTCTT CTACCACATT GTGAAGGAAA	5520
CTTTCAACAT CATGCTTAG TGATTACTAG ATTAATAATCA CCTATCAGAG AATGGTCGAA	5580
GGATATCCAG TGTGTTGACC TTATTATCGG TTTGGCCATT GCAGTATCAC AGGACAAGTC	5640
ATGTATTAAA ACATTGATGA GAAGACTAGC AGATGAATCA TTCATAAATC AATTAAAACA	5700
GTTAACAAAA GAAGAATTAC GGGAGATAAT ATATGGAAAT CAAAGATATT CTTAATGTGA	5760
GTCTGATCCA GACGGATTTA CAGATGCAGA GCAAAGAAGA GGTTTTTGAG GCATTAGCTC	5820
AACTATTGGT TGAGACGGGT TATGTGTCTG ATAGAGACCA ATTTATCGAA GGTCTTTATC	5880
AGAGAGAGGC AGAAGGACAG ACCGGTATTG GGAATTATAT TGCTATTCCC CATAGCAAGA	5940
GTTCTGCTGT GGAGAAGGCG GGGGTAGTCA TAGCTATAAA TCACAATGAG ATTCCTTGGG	6000
AGACCATTGA TGGGAAAGGG GTCAAAGTAA TTGTACTCTT TGCAGTTGGT GATGATACAG	6060
AAGCTGCTAG GGAGCATTTG AAGACCTTAT CACTCTTTGC TCGAAAACCT GGTAATGACG	6120
AAGTTGTTGC CAAATTAGTT CGGGCTCAGA CATCTGATGA TGTGATTGCA GCTTTTGTGT	6180
AATAAGAAAA AATTTTGGAG GGTATCCGTA TGAAAATTGT TGGTGTGCA GCTTGTACTG	6240
TGGGAATTGC CCACACTTAT ATTGCACAGG AAAAATTAGA GAATGCCGCA AAGGTAGCTG	6300

449

GACATGTGAT TCATGTTGAG ACTCAGGGGA CAATAGGGGT AGAAAATGAA TTGACTCAAG	5360
AGCAGATTGA TGCAGCGGAT GTAGTTATTT TAGCAGTTGA TGTAAAGATT TCTGGTATGG	6420
AACGCTTTGA GGGTAAAAAG ATTATCAAGG TTCCAACAGA AGTGGCAGTC AAATCTCCCA	6480
ATAAACTGAT TGCTAAAGCT GTTGAGATTG TTACGAAATA ACTGAAAATA TTAAAGGAGA	6540
AAATATATGT TGAAACACTT AAACTTAAAA GGTCACTTAT TGACAGCCAT TTCCTATATG	6600
ATTCCAATTG TTTGTGGTGC AGGATTCTTA GTTGCCATTG GTTTAGCAAT GGGGGGTGGT	6660
GTTCCTGACG CTCTGTAGC AGGAAAATTC ACTATCTGGG ATGCTTTAGC AACTATGGGT	6720
GGTAAAGCCC TTGGTCTCTT GCCAGTTGTT ATTGCTACAG GTTTGTCTTA CTCGATTGCT	6780
GGTAAGCCAG GGATTGCACC AGGTTTGTGTT GTTGGTCTAA TTGCCAATTC TGTGGTTCA	6840
GGGTTTATCG GTGGTATCTT GGGAGGTTAT ATAGCTGGTT TCTTGGTCA AGCGATTATT	6900
AAAAAGGTCA AAGTACCAA CTGGATTAAA GGTTTAATGC CAACCTTGAT TATTCCTTTT	6960
GTAGCCTCTT TGGTAAGTAG TTTGATTATG ATTTATATTA TTGGAGCGCC TATCGCAGCC	7020
TTTACCAACT GGTGACGAG CTTATTACAA AGCTTGGGAA GTGCTTCAAA TGGTTTGATG	7080
GGGGCAGTTA TTGGAATTCT CAGTGCTGTT GACTTTGGTG GCCCACTTAA TAAACAGTC	7140
TATGCGTTTG TGTGACTTT ACAGGCTGAA GGTGTGAAAG AACCATTGAC TGCTTTACAA	7200
TTGGTGAATA CTGCTACACC AGTTGGATTT GGATTGGCCT ATTTTATCGC GAAATTACTC	7260
AAAAAAAATA TCTATACTCA AGAGGAAATC GAAACATTGA AATCGGCTGT TCCTATGGGG	7320
ATTGTCAATA TTGTGAAGG TGTAATCCG ATTGTTATGA ATAACCTGGT TCCAGGTCTC	7380
ATTGCAACAG GTATCGGTGG TGCTGTTGGT GGTGCTGTTT CTTTGACAAT GGGTGCTGAT	7440
TCTGCTGTGC CATTTGGTGG AGTGCTTATG TTACCAACCA TGACTCGTCC AGTAGCTGGT	7500
ATTTGTGCCT TGTTAGCTAA CATTGTAGTC ACAGGACTTG TCTACGCGAT TTGAAAAAA	7560
CCAATAAAAC ATGCAGAAC AGTTATGACT GTTGAAGAAG AGATTGATTT GTCAGATATT	7620
GAAATTTGT AAGAGGGTAA CGATGTCAAG AATTGAATTT TCACCATCTT TGATGACCAT	7680
GGATTTGGAC AAATTCAAAG AGCAGATTAC TTTTTTGAAT GATAAAGTAG CATCTTATCA	7740
TATCGATATT ATGGATGGCC ATTTTGTTC CAATATTACC TTGTCTCCTT GGTTCATTCA	7800
AGAAGTTCAA AAAATTAGTG ACACACCTTT ATCAGTTCAT CTGATGGTCA CAGACCCAAC	7860
CTTTTGGGTA GATCAAGTTC TCGATTTACA ATGTGAGTAT ATTTGTATTC ATGCTGAAGT	7920
TCTGAATGGT CTGCTTTTC GTTTGATTGA TAAATTCAT GATGCAGGTC TAAAGGCTGG	7980
TGTTGTCTT AATCCTGAAA CACCTGTTTC TACAATCTT CCCTACATTG ATTTACTTGA	8040

450

CAAAGCAACT ATTATGACTG TAGATCCAGG TTTTGCAGGA CAACGCTTTT TGGAGTCTAC	8100
CTTGTATAAA ATCCAAGAAC TCCGTCAGCT TAGAGTTCAG AATGGTTATC ACTACATCAT	8160
TGAGATGGAT GGTTCCTCGA GTCGTAAGAC TTTCAAACAA ATTGATGTGG CAGGACCAGA	8220
TATTTATGTT ATAGGTCGCA GTGGATTATT TGGTTTGGAT GACGATATTG CCAAAGCCTG	8280
GGATATCTGT TCTAGAGATT ACGAAGAAAT GACCGGAAAA ACAATGCCAA TCAAATAATG	8340
GTTTGAGAAG AAATTTATTA GTTAGGAGGA ATATATGTCA CTACAATCAG TTAACGCCAT	8400
TCGTTTCTCT GCGGTAGATG CTATTAACAA ATCTAATTCT GGTCAACCCG GAATTGTCAT	8460
GGGTGCTGCG CCAATGGCTT ATAGCCTATT TACAAAGCAC CTTAGAATPA CACCTGAGCA	8520
GCCAAACTGG ATTAACCGAG ATCGCTTTAT CTGTCTGCG GGTTCATGGAT CAATGCTACT	8580
GTATGCTCTC TTGCATTTAA CAGGTATATA GGATGTATCC ATGGACGAGA TTAATAATTT	8640
CCGGCAATGG GGATCTAAGA CACCTGGTCA TCCTGAAGTG ACGCATACGT CTGGTGTGGA	8700
TGCGACATCT GGTCCGCTTG GTCAGGGGAT TTCTACTGCC GTTGGTTTCG CCAAGCAGA	8760
GCGTTTTTTA GCTGCTAAGT ACAACAAAGA TGGTTTCCCT ATTTTGTACC ATTATACTTA	8820
TGTTATCGCT GGAGACGGTG ACTTCATGGA AGGAGTGTCT GCGGAGGCGG CTTCTTATGC	8880
AGGTCATCAA GCTTTAGATA AGCTTATCGT CCTCTACGAC TCCAACGACA TCTGCTTGGA	8940
TGGTGAGACC AAAGATACTT TCTCTGAAAA TGTTCCGCTC CGTTACGATG CTTATGGTTG	9000
GCATACAGTT CTGGTAGAAG ATGGAACAGA TTTAGCAGCA ATTTCTACAG CAATTGAGAC	9060
GGCCAAGTTT TCTGGTAAAC CGAGTTTGAT TGAAGTGAAG ACGGTAATTG GTTACGGCTC	9120
ACCCAATAAA AGTGGTACAA ATGCTGTTCA TGGTGCACCA CTAGGAGCAG AAGAAACAGG	9180
AGCAACTCGT AAGTTTTTGG GATGGGATTA CGATCCATTT GAAGTACCAG AGGAAGTATA	9240
TTCTGATTTT AAGACAAATG TAGCGGATCG TGGTCAGGAG GCATACGATG CTTGGGCTAG	9300
TTTGGTGTCT GATTACAAGG TTGCTTATCC CGAAGTTGCT AGTGAGATTG ACGCTATTGT	9360
AGCTGAAAA TCCCCTGTAA CCATTACTGA AAAAGACTTC CCTGTCTATG AGAATGGCTT	9420
CTCTCAAGCA ACTCGTAATT CGTCCCAAGA TGCTATTAAT ACAGCAGCAG TTTTACCAAC	9480
CTTCTTAGGT GGATCGGCAG ACTTAGCTCA CTCTAACATG ACCTACATCA AGGCAGATGG	9540
CTTACAAGAT AAATATAATC CATTAAACCG CAATATTCAG TTTGGGGTAC GTGAATTTGC	9600
CATGGGAACA ATCCTCAATG GAATGGCTCT TCATGGTGGT TTACGAGTTT ATGGCGGAAC	9660
CTTCTTTGTT TTCTCTGACT ACGTCAAAGC TGCTATTTCG CTATCAGCCA TTCAGGAGTT	9720
GCCTGTAAC TATGCTTTTA CCCATGATTC AATTGCCGTT GGTGAAGATG GTCCAACTCA	9780
TGAACCAATT GAACATTTGG CAGGTTTACG CTCAATGCCA AACTTGACTG TTATCCGTCC	9840

451

AGCGGATGCC CGTGAAACTC AAGCGGCTTG GCATCATGCC TTGACCAGTA CCACCACTCC	9900
AACTGTCATT GTCTTAACCC GTCAAAACTT GGTAGTTGAA GAAGGGACAG ACTTTGGTAA	9960
GGTCGCTAAA GGAGCCTACG TCGTGTATGA TACCCCGGGA TTTGATACTA TTATCATTCG	10020
TACAGGATCT GAGGTCAATC TAGCTATCAA AGCTGCTAAG GAATTGGTTT TACAAGGTGG	10080
TAAAGTACGT GTGGTATCTA TGCCCTCAAC CGAACTATTT GATGCTCAAG ATGCTACCTA	10140
CAAGGAAGAC ATTTTACCAT CTAAGACTCG TCGTCGTGTG GCCATTGAAA TGGCAGCGAC	10200
CCAAAGTTGG TACAAGTATG TTGGTTTGGG TGGCGCGGTC ATCGGTATTG ACATCTTCGG	10260
TGCGTCTGCC CCAGCTCAGA CTGTGATTGA TAATTATGGA TTTACGGTAG AGAATATCGT	10320
TGCTCAAGTT AAGTCCCTAT AGAAACCAAT TACAATGAAG ATACAGCTGT TGTCAAGACTA	10380
GCAGATGTAG TGATAGACAC TAATCAGATG ATTGTTTATT TAAAACTGT AATGAAAATG	10440
TAATAATTTA TCTACGAAAG TTATAGTAGA TAGTATACAC AATAGAGTAT ACCCTGAAAC	10500
GGTTGCGAAG TACGCTAATC ACTTTGCTAC TGATCTAGAT AGTTTCTTTA ATCAATAAAC	10560
ACAGCATCCA CAGATTGACT TAGGATATTG TAAGTTTTTT GAAAGCTAGA GAGAAGGTCT	10620
CTAAAATTAA AAAACGCATA GTATAGGATG TTGAAATGAT GAACTGCACC CCAAAGTTA	10680
GACAGAAAAA AATCTAACTT TTGGGGTGTT TTTATTATGA AATTAACTTA TGATGATAAA	10740
GTTCACTTCT ATGAACCTAG AAAACAAGGA TATATCTTAG AGAAGCTTTC AAATAAATTT	10800
GGGATAAATA ATTCTAATCT TAGGTACATG ATTAAATTGA TTGATCGTTA CGGAATAGAG	10860
TTCTGCAAAA AAGGGAAAAA TCGTTACTAT TCTCCTGATT TAAAACAAGA AATGATTCAT	10920
AAAGTCTGAC ATGAAGGCTG GACTAAAGAT AGAGTTTCTC TTGAATACGG TCTCCCAAGT	10980
CGTACGATAC TTCTTAACTG GCTAGCACAA TACAGGAAAA ACGGGTATAC TATTGTTGAG	11040
AAAACAAAAG GGAGAGTACC TGAGAGCGGA GAATGCCATC CTAAGAAAGT TAAGAGAACT	11100
CCGATTGAAG GAGGAAAAAG AGAAATAAGA AAGACAGAAA TTGTTCAAGA ATTAATGACT	11160
GAGTTTTCGT TAGATCTTCT TCTAAAAGCC ATTAACTAG CTCGTTGGAC CTACTACTAT	11220
CACCTGAAAC AGCTAGATAA ACCAGATAAG GACCAAGAGC TTAAAGCTGA AATTCATACC	11280
ATCTTTATCG AACACAAGGG AGATTATGCT TATCGCCGGG TTCATTTAGA ACTAAGAAAT	11340
CGTGCTTATC TGGTAAATCA TAAAAGAGTT CAAGGCTTGA TGAAAGTACT CAATTTACAA	11400
GCTAGAATGC GACAGnAACG AAAATATTCT TCTCATAAAG GAG	11443

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5338 base pairs

452

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CCAATTACAT TATATTATCA AAATCGTCGA AACTGGCTCC ATGAATGAGG CAGCCAAGCA	60
ACTCTTTATC ACTCAGCCAA GTCTCTCCAA TGCAGTGCGA GATTTGGAATG	120
CATTGAGATC TTTATCCGCA ATCCCAAGGG AATCACCTTG ACCCGTGATG GCATGGAGTT	180
TCTCTCTTAT GCGCGTCAGG TTGTCGAGCA GACCCAGCTT CTGGAGGAAC GCTATAAAAA	240
TCCTGTGCGC CACCGCGAAC TCTTTAGCGT TTCGTCTCAA CACTATGCCT TTGTGGTCAA	300
TGCCTTTGTC TCTTTGCTCA AGAAAAGCGA TATGGAGAAA TACGAACCTC TCCTTCGTGA	360
AACTCGGACT TGGGAGATTA TCGACGACGT CAAGAAGCTC CGCAGTGAGG TCGGGGTCC	420
CTTCTTAAAC AGTTACAACC GTGATGTTTT AACCAAGATG CTGGATGACA ATCACCTGCT	480
AGCCCACCAT CTCTTCACAG CGCAACCGCA TATCTTTGTC AGCAAGACCA ACCCTCTGGC	540
AAAGAAAGAC AAGGTGAAAC TGTCTGATTT GGAGAAATTC CCTTACCTCA GCTATGACCA	600
AGGGACGCAC AACTCCTTCT ACTTTTCAGA AGAGATTCTT TCTCAAGAAC ACCACAAGAA	660
ATCCATTGTG GTCAGTGACC GTGCCACCCT CTTTAATCTC TTGATTGGTT TGGATGGTTA	720
TACCATTGCG ACAGGGATTT TGAACAGCAA CCTAAACGGA GACAATATCG TTTCTATCCC	780
ACTGGATATT GATGACCCGA TCGAGCTGGT CTATATCCAG CATGAGAAAA CCAGCCTATC	840
TAAGATGGGC GAACGCTTTA TAGACTATCT CCTAGAAGAA GTTCAGTTTG ATAGTTGAGA	900
AATGATAAGA ACCAATATGT AGGCTAGCAA CAACCTGCAC ATTGGTTCTT TTTACTTATA	960
ATTAAAAGTT TCCCCTGCCA ACTTATCAGC TAGCTTGGGA AAGAGAGTAT AAAACTTATG	1020
GGCTAGGTTT AACAAAATCG GGAGATTGAG TTCTCGTTTG TTTTTCCTA TAATCTTGAC	1080
AATCTTTTTA GCCACTGCAT CTGGTCTTAG CAGGAAGCGA TCAACCGATT TAAGATAAGT	1140
TCCATCTGGG TCGGCTTGGT CGAAAAATCC TGTACGGATT GGTCTGGAT TGAATGTTGT	1200
CACATAGACT CCATAGGGCA TAAGTTCGAG TCGCAGAGCA TTTGAAAAAC CAATAGCCGC	1260
AACTTGGTC GCTGAGTAAA GACTAGACTT GCCAGTAGCT ATTAGACCTG CCATGCTGAC	1320
GATGTTGATG ATATGCCCTT TGCTGCTTTC CTTCATACGA GCCGCAAGGT GACGAGACAG	1380
ATTCATCAGG GCAAAGGTAT TGACCTCAA CATCTGGTGA ATATCTTTAT CAGCAATCTG	1440
GTCAAATCCC TCAAAAATCC CGTAACCAGC GTTGTTAATC AAGACATCAA TCTTGCCATA	1500
GCGGAGATAA AGATCAGTTA CCAGAGCTTC TAGGGCTGAA TCGTCGGTAA TATCAATTTC	1560

453

AATCAATTCT GCATGGGAAT AATTTCCGTA GAGTTGGGCT AATTTTTCCT TATTTCTACC	1620
AAGCAAGATG AGTTGGTCAT TGGGCAGGAG TTTGACCATT TCTTGAGCTA GACCACCGCT	1680
AGCTCCGGTA ATGAGAATAG TAGGCATACT TATCCTTTCT GTGACTGCTA GATTTCCACT	1740
TCTTCCAAGT CTTTGACCAC ATGGACATTT TCAAAAATTG TGGCAGCGTC TTTCTTGAGT	1800
TTGCTAATAT CTTTGAGAG GAAACGGGCA CTGATATGGT TGAGTAGGAG GCGTTTGGA	1860
CCTGCTTCTA CCGCTACTTG TGCAGCTTGC ATATTAGTTG AGTGACCATG GTTACGAGCA	1920
ATTTTTCAT CACCCTTGCC ATAAGTGGAC TCATGAACTA GGACATCTGC ATTGACAGCC	1980
AGACGCACAC TGGCACCCTG TTTTCGAGTG TCTCCTAAAA TAGTGATAAT CTTACCTGGA	2040
CGTGGCGCTG AGATATAGTC TGCTGCCTTG ATTTTCAGTTC CGTCTTCCAA AACAAGATCC	2100
TGGCCGTTTT TGATTTTACC AAAAAGCGGG CCGAACGGAA CACCAGCAGC CTTGAGTTTT	2160
TCAGCATCCA GCGTCCCTTC TAGATCCTTT TGCATGACAC GATAGCCAAC ACAGAAAATA	2220
GTGTGGTCCA GCTCCTCTGC ATACACAGTG AATTTATCGG TTTCAAGAAT TTTACCCAGA	2280
GAATCTTGGT CAAACTCATG GAAATGAATG CGGTAGGGCA GACGAGAAC TGACACACGA	2340
AGGCTGGTTA AGACAAATGA CTTGATTCCT TGAGGTCCCT AGATTTCCTA ATCTGTCTGC	2400
TCTTCATTGG CTGAAAGGC ACGGCTAGAA AGGAAACCTG GCAAACCAA AATGTGGTCT	2460
CCATGCAGAT GGGTAATAAA GATTTTGCTG ACCTTACGTG GTCGAATTGT GGTTCACAGA	2520
ATGCGATTTT GCGTACCTTC TCCACAGTCA AAGAGCCAAA CTTCGTTAAT CTCATCCAAA	2580
AGTTTCAGGG CGAGACTTGA AACGTTGCGG GCTTTAGAGG GCTGACCAGC CCCCCTTCCT	2640
AAAAATTGAA TATCCATTTC ATACTTTCTA ATTAATCAAT ATATAACATG GCTGTGCGGT	2700
TTTCCGATCG GAAATAGCGT TTGCCAGAAA AAGCAGCAGC TTCTTGCAAT AAATCCTCTT	2760
GGCTGTAGCC TTTGAGACGT TTTGACCAT CAGCCAATCT TTCCAAATCA GTCAAAGCTG	2820
TGAGACTTTC TAGGCTGATA ACTTCCTCGT CCTCGACAGG CTTCATGTAA ATCTTACCAG	2880
ACTCTTCAA GACTAATTGA TGGGGGAAAA TTTCGCAAT TTCAAAGAGC AAGTCATCCG	2940
AGATTTTCTC CTCATTTTCA AAGAAAATCC GACCAAGGCC GTCACCTCTA TAACAAAAAC	3000
CAAAGGATTT ACCAGACAGA TTAAGCCGAA TAAAAGGCTT ATTTTCTAGG GTGAAACTTG	3060
GCTCAGTATT GTAAAGATTC AGTTCTGAC TGAGTTCTGC AAAATAATCC GTCGCAGCCT	3120
GAGGACTCTT TTTCTGATAG AGTTCTGCAA AGTAGGCATT AACAACACTT GCGGAGGTG	3180
TAATAAGTGT TAACTGCTCC TGATCTGTTT TACCAGCTAG AAGCTGATCC AGATAGACCT	3240
TGTCCAGACT TGTATAACCT CCATACTTTA GAGCCAAAGT TTTAATATCA GTCATAAAAT	3300

454

TCTTCTAACC	TCCATTTATT	TTTCTCGGAA	ATGTAGCCTG	TAATCACTTC	GCCGTCTTCC	3360
TGATAATCAC	GTTCTTCCAG	AATTGCAACA	CTCTCTAAAT	CATGAATCTT	GTAGGACTTT	3420
GAAAAAGGCA	CTCGCAGGGT	AAATGCTTCA	AAAATTTCTT	TAATCTTATC	TAGCAATAAT	3480
GCTTGCAAGT	TTTCACGACT	GTCCTCAGAC	TTGGCAGAAA	TGAGGGTATA	TGGCGTTTGG	3540
GTAGGCGTGA	AATCCTCCAC	CAAATCCGCT	TTATTATAAA	GCGTCAAGTG	AGGAATATCT	3600
TCCATGTCCA	GGTCTTTCAT	GATGGAGAGA	ACCGTTTTTT	CATGCTCCTC	GTGGTAAGGA	3660
TTGCTAGCAT	CGATAACATG	AACCAGAAGG	TCCACATGCT	TGCTTCTTTC	CAAGGTTGAC	3720
TTGAAACTGG	ACACCAACTC	TGTCGGCAAA	TCTTGGATAA	AGCCAACGGT	ATCTGTCAAA	3780
GTTACTTGGA	GATTGCCTCC	CAGATGAATA	CTCTTGGTTG	TCGCATCCAG	AGTCGCAAAG	3840
AGCTCATCTG	CTTCATACTG	GGTCTTACTG	GTCAAGATGT	TCATGATAGT	TGATTTCCCA	3900
GCATTAGTAT	AACCAATCAA	ACCAATCTTA	AAAGTGCTAG	ACTCCAAACG	TTTTTCTCTG	3960
ACAGTCGCAC	GATTTTTCTC	AACCACCTTG	AGCTGGCGCT	CGATATCCGT	GATTTGATTG	4020
CGAACGCTAC	GACGGTTCAG	CTCCAGCTGG	CTTTCACCAG	GACCACGGGA	ACCAATTCCC	4080
CCTGCTGAC	GGCTGAGCAT	AATCCCTGA	CCAACCAAGC	GAGGCAAAAG	GTATTTGAGT	4140
TGGGCTAGGT	GGACTTGAG	CTTCCCTTCA	TGGCTTCGAG	CCCGCATGGC	AAAGATATCC	4200
AAAATCAACT	GCATACGGTC	AATGACCTTA	ACACCGAGAA	CTTCCTCTAG	ATTGACATTC	4260
TGCCTTGGGG	TCAGACGATT	GTTGACGATG	ACAGTAGTGA	TTTCTTCTGC	ATCCACCATA	4320
AGCGCAATCT	CTTCCAACCT	ACCAGAGCCG	ACGAAGGTCT	TGGAATCATA	TTTTTCACGT	4380
TTTTGTCTGT	AGCTATCTAC	AACGACTGCC	CCTGCCGTTT	TCGCTAAACT	AGCCAATTCT	4440
TCCATGGAGA	GGTCAAAACT	GTCCATACCC	TGCAATTCCA	CACCAATCAG	CAGGACTCGC	4500
TCCTCTTTT	TCTCCGTTTC	AATCATCTAA	AAACTCCTCT	ATCTGGCTTA	AAATGCGGTC	4560
TTGTACACCA	GATTCTCCAA	TCTGATAAAA	GGTGACCTGC	ATGCGATTAC	GGAACCAGGT	4620
CAGCTGACGC	TTGGCAAAAC	GACGAGTCGC	CTGTTTAAGA	CTCTCACTAG	CTTCTCCAA	4680
GGTCTGCTCT	CCACGGAAAT	AAGGAAAGAG	TTCCCTTATAG	CCAATTCCTT	TAGCAGCCTG	4740
TACATTAGGG	GAATGGTCAA	ACAGCCACTT	GGCCTCATCC	AAAAGCCCAG	CCTCAAACAT	4800
CAAATCCACT	CGGTGGTTGA	TACGCTCATA	AAGTTGACTA	CGTTCATCAT	CCAAGCAGAT	4860
AATCAGCGGT	TCATACAAGG	TCTCTTGATT	TTCCAAATCC	TGACCAAAAT	GGGCAATTTT	4920
TAAGGCACGC	ATAGCACGAC	GACGATTAAA	CTGGGGAATC	TCAAGGCCTG	CTTGATCCAC	4980
CAAATGGGCT	AATTCCTCAT	CTGAATATGG	CTCCAAACTA	GCTCGATAAG	CTAAATCTC	5040
CTCATGAGGA	GTCTCCCCAC	CTAGGTGGTA	ACCTTCTAGC	AAGCTCTGGA	TATAAAGTCC	5100

455

AGTCCCACCG GCGATAATGG CTAGCTTGCC ACGGTTGTGA ATACCCTCAA TAGTCATCTT 5160
 AGCTTCTGAA ACAAATCAA AAGCCGAGTA AGACTCGGTT ATCTCTCTAA CATCGATTAA 5220
 ATGATGAGGA ACAGCTGCCT GCTCTTCTGG ACTAGCCTTG GCCGTCCCAA TATCAAGTCC 5280
 TCGATAGACT TGCTGGCTAT CTCCACTAAC CACTTCGCCA TTAACGCT TTGCGGGG 5338

(2) INFORMATION FOR SEQ ID NO: 51:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19446 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

CGGAAACCCA TCTAGTCTCC ATCGTTTGGG AGACCAAGCA ACACGAATCT TAGATGCTTC 60
 TCGCCAACAG ATTGCAGATT TAATCGGTAA GAAAGCGAT GAAATCTTCT TTACCTCGGG 120
 TGGAACAGAA GGGGATAACT GGCTTATCAA GGTGTGGCC TTTGAAAAAG CTCAGTTTGG 180
 CAAGCACATC ATTGTTTCAG CCATTGAACA TCCAGCAGTC AAAGAGTCAG CCCTCTGGTT 240
 GAAAAGTCAA GGATTGGAAG TGGATTTTGC TCCAGTTGAT AAGAAAGGCT TGGTCGATGT 300
 TGAGGCGTTA CAGTTTGTAT ACGGCATGAT ACAATCCTCG TTTCCATCAT GGCTGTGAAC 360
 AATGAAATCG GCTCTATCCA ACCTATTGAG GCTATTTTTCAG AATTCTTGGC AGACAAGCCG 420
 ACTATTTCTT TCCACGTGTA TGCGGTTTTCAG GCGCTTGCCA AAATTCGGAC TGAAAAGTAT 480
 CTGACAGAAC GGGTGGATTG CGCGACTTTC TCTAGTCACA AGTTCCACGG GGTTCGAGGT 540
 GTTGGCTTTG TCTATATCAA ATCTGGCAAG AAGATTACAC CTCTTCTTAC AGGTGGTGGC 600
 CAGGAGCGAG ATTATCGTTC GACAACTGAA AATGTGGCAG GGATTGCAGC GACAGCCAAG 660
 GCCCTCCGTT TGTCTATGGA AAAGCTAGAT ATCTTTAGGA GCAAGACTGG GCAGATGAAG 720
 GCAGTGATTC GCCAAGCTCT TCTGAACTAT CCGGATATTT TTGTCTTTTC AGATGAGGAA 780
 AACTTTGCAC CTCATATTCT GACTTTTGGG ATCAAAGGTG TTCGAGGTGA AGTCATCGTT 840
 CACGCCTTTG AAGACTATGA TATTTTCATC TCAACAACCT CAGCTTGTTT ATCTAAGGCA 900
 GGAAAACCCG CCGGTACCTT GATTGCCATG GGAGTGGACA AAGATAAGGC CAAGTCAGCT 960
 GTGCGTCTTA GCCTAGACTT GGAAAATGAT ATGAGTCAGG TCGAGCAGTT TTTGACCAAG 1020
 TTAAAATTGA TTTACAATCA AACTAGAAAA GTAAGATAGG AGCATTCATG CAGTATTCAG 1080
 AAATTATGAT TCGCTACGGA GAGTTGTCAA CCAAGGGTAA AAACCGTATG CGTTTCATCA 1140

456					
ATAAACTTCG	TAATAATATT	TCGGACGTTT	TGTCTATCTA	TACCCAAGTT	AAGGTAACAG 1200
CAGATCGCGA	CCGTGCCCAC	GCTTACCTCA	ATGGAGCTGA	TTACACAGCA	GTTGCAGAAT 1260
CTCTCAAACA	AGTTTTTGGG	ATTCAAAACT	TTTCTCCTGT	TTATAAGGTT	GAAAAATCTG 1320
TAGAAGTTTT	GAAGTCTTCT	GTCCAAGAGA	TTATGCGGGA	CATCTACAAG	GAAGGTATGA 1380
CCTTTAAGAT	TTCTAGCAAG	CGTAGCGACC	ACAACTTTGA	ACTTGATAGT	CGTGAACCTCA 1440
ACCAAACACT	TGGAGGGGCT	GTATTCGAAG	CCATTCCAAA	TGTGCAAGTT	CAAATGAAAA 1500
GTCCTGACAT	CAATCTTCAG	GTGGAGATTC	GTGAAGAAGC	AGCCTATCTT	TCTTATGAAA 1560
CCATTCGTGG	GGCTGGTGGT	TTGCCAGTTG	GAACCTCAGG	TAAAGGGATG	CTCATGTTGT 1620
CAGGAGGGAT	TGACTCACCT	GTAGCAGGTT	ATCTTGCTCT	TAAGCGTGGG	GTGGATATCG 1680
AGGCAGTTCA	CTTTGCTAGT	CCACCATATA	CTAGTCCTGG	TGCCCTCAAG	AAAGCGCAGG 1740
ACTTGACCCG	TAAATTGACC	AAGTTTGCGG	GAAATATCCA	GTTTATAGAG	GTGCCTTTCA 1800
CAGAGATTCA	AGAGGAAATC	AAAGCCAAAG	CGCCAGAAGC	TTATTTGATG	ACTCTAACTC 1860
GTCGCTTTAT	GATGCGGATT	ACTGACCGTA	TTCGTGAGGT	ACGAAATGGT	TTGGTTATCA 1920
TCAATGGGGA	AAGTCTAGGT	CAAGTAGCCA	GCCAAACCCT	TGAAAGTATG	AAGGCTATCA 1980
ATGCTGTTAC	CAACACTCCC	ATCATTCGTC	CTGTGGTTAC	CATGGACAAG	TTGGAAATCA 2040
TTGACATCGC	CCAGGAAATC	GATACCTTTG	ACATTTCAAT	CCAACCGTTT	GAAGACTGTT 2100
GTACCATTTT	TGCACCAGAT	CGTCCAAAAA	CAATCCTAA	AATTAAGAAT	GCGGAGCAGT 2160
ACGAAGCGCG	TATGGATGTT	GAAGGCTTGG	TTGAGCGAGC	AGTGGCTGGA	ATCATGATTA 2220
CTGAAATCAC	ACCTCAAGCC	GAAAAAGATG	AAGTTGATGA	CTTGATTGAC	AATCTGCTCT 2280
AATTCAGAAA	ATCCAAAAGA	ATAGCGAAAA	TCAGTAAAAA	AAGTTAGTTT	TTTCTCTAAA 2340
AACAGGTAAT	AACTAATCTT	TTTTTATTTT	TATGATATAA	TGATATAAAA	TTTGAATAT 2400
AGAGAGTTTT	CTGACAATGA	ATCAATCCTA	CTTTTATCTA	AAAATGAAAG	AACACAAACT 2460
CAAGGTTCCCT	TATACAGGTA	AGGAGCGCCG	TGTACGTATT	CTTCTTCCTA	AAGATTATGA 2520
GAAAGATACA	GACCGTTCCT	ATCCTGTTGT	ATACTTTCAT	GACGGGCAAA	ATGTTTTTAA 2580
TAGCAAAGAG	TCTTTCATTG	GACATTCATG	GAAGATTATC	CCAGCTATCA	AACGAAATCC 2640
GGATATCAGT	CGCATGATTG	TCGTTGCTAT	TGACAATGAT	GGTATGGGGC	GGATGAATGA 2700
GTATGCGGCT	TGGAAGTTCC	AAGAATCTCC	TATCCCAGGG	CAGCAGTTTG	GTGGTAAGGG 2760
TGTGGAGTAT	GCTGAGTTTG	TCATGGAGGT	GGTCAAGCCT	TTTATCGATG	AGACCTATCG 2820
TACAAAAGCA	GACTGCCAGC	ATACGGCTAT	GATTGGTTCC	TCACTAGGAG	GCAATATTAC 2880
CCAGTTTATC	GGTTTGGAAT	ACCAAGACCA	AATTGGTTGC	TTGGGCGTTT	TTTCATCTGC 2940

457

AAACTGGCTC CACCAAGAAG CCTTTAACCG CTATTTTCGAG TGCCAGAAAC TATCGCCTGA	3000
CCAGCGCATC TTCATCTATG TAGGAACAGA AGAAGCAGAT GATACAGACA AGACCTTGAT	3060
GGATGGCAAT ATCAAACAAG CCTATATCGA CTCGTCGCTT TGCTATTACC ATGATTTGAT	3120
AGCAGGGGGA GTACATCTGG ATAATCTTGT GCTAAAAGTT CAGTCTGGTG CCATCCATAG	3180
TGAAATCCCT TGGTCAGAAA ATCTACCAGA TTGTCTGAGA TTTTTCGAG AAAAATGGTA	3240
AGTTAAGAAA GGAAAAACG AAATGCATAT TGAACATCTT AGCCACTGGA GTGGTCATCT	3300
TAACCGTGAA ATGTACCTTA ACCGTTATGG ACATGGTGGG ATTCCAGTTG TGGTCTTTGC	3360
TTCATCAGGT GGTAGTCACA ACGAATACTA TGATTTTGGC ATGATTGATG CCTGTGCTTC	3420
CTTTATCGAG GAAGGCCTTG TCCAGTTCTT TACCCTATCT AGTTTGGATA GTGAGAGCTG	3480
GTTGGCTACT TGGAAAAATG CTCATGACCA AGCGGAAATG CACCGTGCCT ACGAACGTTA	3540
TGTGATTGAG GAGGCCATTC TTTTATCAAG CACAAGACAG GTTGGTTTGA TGGCATGATG	3600
ACGACAGGTT GCTCTATGGG AGCCTATCAT GCACTCAATT TCTTCCTCCA GCATCCAGAT	3660
GTCTTTACCA AAGTGATTGC TCTCAGTGGT GTTTACGACG CACGTTTCTT TGTGCGTGAT	3720
TACTACAACG ATGATGCTAT TTACCAAAAC TCGCCAGTAG ATTATATTTG GAACCAAAAC	3780
GACGGCTGGT TTATTGACCG TTACCGTCAG GCAGAGATTG TGCTGTGTAC GGGGCTTGGA	3840
GCCTGGGAAC AAGATGGTTT GCCATCCTTT TACAAGCTCA AAGAAGCCTT TGACAAGAAA	3900
CAAATTCCAG CCTGGTTTGC TGAATGGGGA CATGATGTCG CCCATGACTG GGAATGGTGG	3960
CGTAAACAAA TGCCTTATTT CCTCGGTAAT CTCTATTTAT AAAAGGAGTT ACCTATGAAT	4020
TACCTTGTTA TTTCTCCCTA CTATCCACAA AACTTTCAAC AGTTTACCAT CGAACTAGCT	4080
AATAAAGGCA TCACAGTCTT GGAATTTGGT CAAGAGTCTT ACGAGCAATT GGATGAGCCC	4140
TTGCGCAATA GCTTGACCGA GTATTTTCGT GTTGATAATC TTGAGAACAT AGATGAAGTC	4200
AAACGTGCAG TTGCTTTTCT CTTTATAAAA CATGGTCCAA TTGGCCGCAT CGAGTCTCAC	4260
AATGAATACT GGCTTGAGCT AGACGCAACA CTCAGAGAAC AATTCAATGT TTTTGGTGCC	4320
AAACCAGAGG ATCTCAAAA GACGAAATAT AAGTCTGAAA TGAAGAACT TTTCAAAAAA	4380
GCAGGTGTTC CTGTGGTACC TGGAGCTGTT ATCAAGACGG AAGCAGATGT TGATCAAGCA	4440
GTGAAAGAAA TCGGTCTTCC AATGATTGCC AAACCTGATA ATGGAGTGGG AGCAGCCGCA	4500
ACCTTTAAAC TTGAGACAGA AGACGATATC AATCACTTCA AGCAAGAATG GGACCATTCA	4560
ACCCTTTATT TCTTTGAAAA ATTTGTCACT TCCAGCGAAA TCTGTACCTT TGACGGGCTC	4620
GTGGACAAGG ATGGAAAGAT TGTCTTCTCA ACAACCTTTG ACTACGCCTA TACACCGCTT	4680

458

GACCTCATGA TTTATAAGAT GGACAATTCT TATTATGTGC TCAAGGATAT GGATCCTAAA	4740
CTGCGCAAGT ATGGGGAAGC AATTGTCAAA GAATTTGGTA TGAAAGAACG GTTTTTCAT	4800
ATTGAGTTCT TCCGTGAGGG GGACGATTAT ATTACCATCG AGTACAATAA CCGCCCTGCA	4860
GGTGGTTTTA CCATTGATGT TTATAACTTT GCTCATTCCT TGGACCTTTA TCGTGGCTAT	4920
GCAGCTATTG TCGCAGGAGA GGAGTTCCTG GCGTCAGACT TTGAAACTCA GTATTGTTTG	4980
GCTACTTCTC GCCGTGCAAA TGCTCACTAT GTTTATTCAG AAGAGGATTT GCTTGCCAAA	5040
TATAGCCAGC AGTTCAAGGT TAAAAAAGTC ATGCCAGCTG CCTTCGCGGA ACTTCAAGGA	5100
GATTACCTGT ATATGCTGAC CACTCCGAGT CGACAAGAAA TGGAGCAGAT GATTGCAGAT	5160
TTGCGACAAC GTCAAGAATA AGAACTATCG GATTAAGGAA ATTAACCTCC TTAATCCTTT	5220
TGTTTTGTCT GATAAAAAAT AAGAGCATCC CAACAAGGTA GCTATCATAA AACTTGTTCCG	5280
ATAACTATTT GAAGCAGGAT TAGGTGGTCA GAAATTAAAT TTTAATATTT CAATTGAGTC	5340
ATAGTATTGT GTTGCGTAT CCTTAAATCA GCTAAAAGGA TCCATGACGA CACCTATACG	5400
ATATAGTTTT CAAGATACCA AACAAGTCTA TTAATATTCA ATGAAAATCA AAGAGCAAAC	5460
TAGGAAGCTA GCCGCAGGTT TCTCAAAACA CTGTTTTGAG GTTGTGGATA GAACTGACAG	5520
AGTCAGTATC ATATACTACG GCAAGGTGAA GCTGACGTGG TTTGAAGAGA TTTTCGAAGA	5580
GTATAAAATA TTCAGGTGAC GCATAGATAT AGTTAATTGA AGCTTTGTTT GAAATCTGAT	5640
AAAATAATGA TATTACTAAG TTTTAAAAAC TAAAGAAAAG GGAAGATATG ATTACAGGCG	5700
AATTAAAAAA TAAATCGAT CAGCTGTGGG AAATTCCTTG GACAGAAGGA AACGCAAATC	5760
CTTTAACAAA TATTGAACAG TTGACTTATC TCTTATTTAT GAAAGATTG GATAGTGTCG	5820
AGCTTGGACG TGAAAGTGAT GCTGAATTC TAGGGATTCC TTATGAGGGA GTTTTCCAA	5880
AAGATAAACC TGAATACCGT TGGTCAACTT TTAATAATAT AGGAGATGCT CAGGAAGTTT	5940
ATCGTTTAAT GACTCAGGAG ATTTTCCGT TTATTAAAA TCTCAAGGGG GATACAGATG	6000
ATACAGCCTT TTCACGATAT ATGCCAGAAG CTATTTTCA AATAAATAAA CCTGCTACGC	6060
TTCAAAAGGC AATTCTATC TTAGATGTTT TTCCAAC TAG GGAATTAGAT GTAGATTTTG	6120
ATAATGACAA ACAAAGTATT ACTGATATCG GAGATATCTA TGAATATCTG TTATCAAAAT	6180
TGTCGACCGC AGGTAAAAAT GGACAGTTCC GTACACCTCG TCACATCATC GATATGATGG	6240
TTGAGTTGAT GCAACCGACT ATCAAAGATA TCATCTCAGA TCCCCTATG GGTTCTGCTG	6300
GCTTCTTAGT ATCTGCTAGC CGTTACTTAA AGCGTAAGAA AGATGAATGG GAAACCAATA	6360
CAGATAATAT CAATCATTTT CATAATCAGA TGTTTCATGG AAATGATACG GATACGACTA	6420
TGTTGAGACT TGGGGCGATG AACATGATGC TACATGGAGT AGAAAATCCA CAAATCAGTT	6480

ACCTTGACTC GCTGTCTCAA GATAATGAAG AAGCCGATAA ATATACTTTG GTTTTAGCAA 6540
ATCCTCCTTT TAAGGGCTCA CTTGACTACA ATTCAACCTC TAATGACCTT CTTGCAACCG 6600
TAAAAACCAA AAAACAGAA TTACTCTTTC TTTCTCTTTT CTTGCGAACT TTAACCAG 6660
GTGGACGAGC AGCAGTTATC GTACCTGATG GTGTCCTTTT TGGTTCGTCT AAAGCTCATA 6720
AAGGAATTCG TCAGGAAATT GTAGAGAATC ATAAGCTTGA TGCTGTAATC TCAATGCCTA 6780
GTGGTGTGTT CAAGCCTTAT GCTGGAGTTT CAACTGCCAT TCTCATCTTT AAAAAAAGTG 6840
GTAATGGTGG TACTGACAAA GTCTGGTTTT ACGATATGAA AGCGGATGGT TTAAGTTTGG 6900
ATGATAAGCG ACAACCGATT AGCGACAATG ATATTCCAGA TATTATCGAA CGCTTTCATC 6960
ATCTTGAAAA AGAAGCAGAA CGTCAGAGAA CGGATCAATC TTTCTTTGTT CCAGTTGCTG 7020
AGATAAAGGA AAATGATTAT GATTTGTCTA TCAATAAATA TAAAGAGATT GAGTATGAAA 7080
AAGTTGAGTA TGAACCAACA GAAGTCATAT TAAAGAAAAT CAATGATTTA GAAAAAGAAA 7140
TTCAAGCTGG CTTGGCTGAA TTGGAAAAAT TACTCAAGTA GGGAGGTGGC TGTATGAAAA 7200
AAGTGAAGTT GGGGAAGTC TTATCTCTAA AAAAAGGCAA GAAAGCCACT GTACTTGCTG 7260
AACAAACAAC TCTAAGCCAA CGTTATATTC AAATAGATGA TTAAGAAAAT AATAATAATT 7320
TAAATTCAC TGAAAGTTTA AATATGACTG AAGCACTCCC AGATGATATT CTGATAGCAT 7380
GGGATGGAGC TAATGCAGGA ACAGTTGGTT ATGGATTATC GGGAGCTGTT GGTAGTACAA 7440
TTACGGTCTT AAAAAAGAAT GAGCGATACA AAGAAAAAAT TATATCAGAT TACTTGGGAG 7500
TCTTTTGGGA AAGTAAATCG CAGTATTTAC GAGATCATTC AACAGGTGCA ACAATTCCTC 7560
ATTTAAACAA GAATATATTA CTTGATTTAC AATTAGAATT GCTAGGTATC GAAGAACAAG 7620
AGAACATTAT CTGTATTCTT AATACGATTA AAAGGCTTAT TACTAAAAGA AAATTTAGT 7680
TAGATGAACT AAAGTTGCTC GTCAAATCCC GATTTAACGA GATGTTTGGG GAAAATAAAA 7740
TATTTGAAAG CATTGATAAC TTATTTGATA TTATAGATGG TGATAGGGGC AAAAATTATC 7800
CTAAATCAGA TGAGTTGTTT AGTGAGGAGT ACTGTTTATT TTAAATACA AAGAATGTTA 7860
CTAAAAACGG ATTTTCATTC GATACAAAGC AATTTATCAC TAAACAAAAG GATAAATTAC 7920
TTCGAAAAGG CAAACTTGAG CGTTATGATA TAGTCTTGAC AACAGAGGT ACTGTTGGAA 7980
ATGTAGCGTA CTACGATGAA TTAATAAAAT ATAAACATTT ACGTATAAAT TCAGGTATGG 8040
TAATATTACG TCCAAGACA CCAATCTAA ATCAGAAATT TATTATCCAT GTTTTAAGGA 8100
ATAATAATTA TAGTCGAGTG ATATCAGGAA GTGCTCAGCC TCAGTTACCA ATTACAAAAT 8160
TAAAAAAAT ACTTCTCCCC CTCCCCCAC TAGCCCTCCA AAATGAGTTC GCAGACTTTG 8220

460

TAGTCCAGGT	CGACAAATCA	CAATTGCTT	GTGAGATAGC	TATAAAAGTG	TGGAGAAATA	8280
GCTTGAAATT	TAGTATAATA	TAGCTAACT	ATTTGTTTAA	AGTGAGAAAA	AAATGGGAAA	8340
TTTTAGCTTT	CTTTTAAAA	ATGACGAATA	TGAATCTTTT	TCAAAACCTT	GCATTGAAGC	8400
TGAGAATATG	ATTGCTACAT	CAACTGTGGC	TACTGCCTTT	ATGGCGCGTC	GTGCTTTAGA	8460
GCAGGCTGTC	CATTGGATAT	ATAGTCACGA	TTCATATTTA	GAAGCTCCCT	ATCGTGCTAC	8520
TCTATCTTCT	TAGTATGGG	ATGATGATTT	TAGGGATATC	GTAGATTCTG	AACTCCACAA	8580
GCAGATAGTT	CTGTTGATTG	GGTGGGAAA	CCATGCTGCT	CATGGTGGTG	AAATTAAGGA	8640
ACGAGAAGCG	ATTTTAGCTT	TGCATCATTT	GTATCAGTTT	GTTAATTTTA	TCGATTATTG	8700
TTACAGCAAT	GAGTFTGTGG	AGCGTTATTT	TGATGAGAAG	TGCTTACCAC	TTTCAGCAAA	8760
CATCAAAATAC	CGAGAAACTC	CACAATCTAT	GATAAAGTTA	CAAGACAGTT	TACCAGAACT	8820
GCCTGATTTT	CATGAACAGA	TGGCTGCTCA	GTCCGTAGAA	GTTCAAGAGA	CTTATACTGA	8880
AAAACGTGAG	ACTGCAGCGC	AACGGCAAGA	TGTGCCTTTC	CATATTGATC	AATTATCTGA	8940
GGCAGAGACA	AGAAAGCTCT	TTATTGATAT	CGATCTCCGT	TTAGCAGGAT	GGATATTTGA	9000
AGAAAAGTGT	CGTGTGAGA	TAGCCGTTGA	TGGTCTCAAG	CACGGTTCAG	GAATTGGTTA	9060
CTGTGACTAT	GTACTTTATG	GTAAAAATGG	GAAAATTTTA	GCGATTGTGG	AGGCTAAAAA	9120
AGCCTCTGTC	AATCCAGAAG	TAGGGGAAGT	ACAGGTCAAA	GAATATGCTG	AAGCTTTGGA	9180
GAAACATATC	GGCTATCAGC	CAATTGCTT	TATTACAAAT	GGGTGAAGC	ACTATATACT	9240
TGATGGTCCG	AACCGCCGCC	AGATTGCAGG	CTTTTACTCT	CAAGAAGAAT	TGCAATTAGT	9300
GATGGATAGA	CGTCATCTTC	AAAAACCGCT	TGAGGATATT	TCTAGTAAAA	TTAGGGACGA	9360
TATTTCCGGG	CGTCACTACC	AAAAACATGC	CATTGCAAGC	GTTTGTGAAG	CTTTCTCTGA	9420
TCATCGTAGA	CAGGCACTTT	TGGTTATGGC	AACTGGGGCG	GGGAAAACTC	GTACAGCAGT	9480
TTCTCTAGTT	GATATCTTAT	CACGTCATAA	CTGGGTAAAA	AACCTTCTCT	TCTTAGCCGA	9540
TAGAACTTCC	TTGGTTAAGC	AAGCATATGA	TTTCGTTAGA	AAATTACTCC	CAGATCTTTC	9600
CGTTTGTAAC	TTCTTAGAAG	ATAAAGAAGG	AGCTCAATCA	AGTCGCATGG	TCTTTTCAAC	9660
TTATCCGACC	ATGATTGGAG	CGATTAGTGG	TCAAGAAGAA	GTAAATCAAC	GCCCTTTCAC	9720
TGTTGGGCAT	TTTGACCTTA	TCATAATTGA	CGAATCTCAC	CGTTCTATTT	ATCAGAAATA	9780
CAAGTCCATT	TTTGATTATT	TTGATGCAAG	AATTGTAGGC	TTAACAGCTA	CTCCGCGTCA	9840
AGATTTAGAT	AAAAACACCT	ATGGATTCTT	TAATTGGAG	AATGGGGTTC	CAACATATGC	9900
ATATGATTTG	GAAGAGGCTG	TTAAAGACGG	ATATTTAGTA	GCCTATCATT	CTATCGAAAC	9960
CAAACTGAAA	CTACCTACGG	ATGGTCTACA	TTATGATGAT	TTGTCCGAAG	AAGAAAAGGA	10020

461

ACATTTTGAT AGCAAATTG AAGACAATAG CTGTGAAAAA GATATTGATG GGAGTGTATT	10080
TAATTCCTTT ATTTTCAATA AAAGTACAGT AGAAATTGTT TTAAATGAAC TCATGACAAG	10140
AGGAATTCAG ACAGCCTCGG GTGATGAAAT TGGTAAAACT ATTATTTTGT CTA AAAATCA	10200
TGATCATGCG GAATATATCA GAGGTATTTT TAACAACCGC TATCCTGAAA AAGGGAGCGA	10260
CTATGCTCAG GTGATTGATT ATAGTATTAA GCATTATCAG ACCTTGATTG ATGATTTTAA	10320
AATTAAGGAG AAGTATCCTC AAATTGCGAT TTCTGTCGAT ATGTTAGATA CAGGTATTGA	10380
TGTACCAGAG GTTGTTAATT TAGTCTTCTT CAAGAAAGTA CGCTCTAAAA CTAAGTTTGT	10440
GCAGATGATT GGTGAGGAA CCCGTCTATG TAAAGATTTA TTTGGACCTG AGCAGGATAA	10500
GGAAAACTTC TTGGTATTTG ATTATGGGGA CAATTTTGAT TATTTTCGTG CAGATCCAAG	10560
AGATGGAGAG GGTGTCACA TTGTTTCGCT GACTCAGCGT TTATTTAATA TCAAAGTGGA	10620
CTTGATTCGA GAACTTCAGG GACTCCAATA CCAAGAAGAT CAGTTTGCGA GAGCATACCG	10680
TCAGCAGCTT GTCTCGGAAC TTCAAGGTCG TATAGAGAGC TTAAATGAGT TGGACTTCAG	10740
GGTTCGTATG GTTTTAGATA CAGTTTATAG CTATAGGAAA TTGGAAGTT GGCAGAATCT	10800
AACTGCTGTT ACAAGTGAAA CCATTCAAAA AAATCTCTCT CCGCTTTTAT TTGATGAAGA	10860
TAAAGAAGAT GAGATGGCGA GGAGATTGA TTTGTGGTTG CTTCATATTC AGTTGGGGCA	10920
ACTGACAGCT AAATCTTCCA CTGTTTCATAT TTCCCAAGTG ATGAAGACGG CTAGAGCTCT	10980
TTCTGCTATT GGCAATATCC CGCAGGTTTT TGAGCAGGCT GAAATTATCA GGAAAGTACA	11040
GGAGCCTGAA TTTTGGAAG AAGTTAACTT GTCTGATTG GAAAAATTC GTCTTGCTAT	11100
TCGAGATTTA TTACAGTTTT TGGATAAAAC AGACCGTAAA CCCTACTATG TTAACTTTGA	11160
AGATCGTATA CTCTCCACTG TTCACGAGAC CACAGCATTT TTGCAGGTCA ACGATCTTCG	11220
GTCTTACAAT GAAAAAGTTG AGCATTATTT GAAAACTCAT CTGGATGAGG AGTCCATTTT	11280
TAAGCTATAC CATAATAAAA AGTTGACATC TGATGATATG CTTGCACTTG AAAAATTGCT	11340
TTGGGAAAAA TTAGGTAGTA AAGCAGACTA CCAAAGTCAT TATGAAAATA AGGCAATTCC	11400
GAGATTGGTT CGTGAGATTA TTGGCTTAGA TAGAGAGTCT GCCAATCGTA TTTTCTTAA	11460
ATTTTGTGCG GATGAGAATC TTAATGCCAG GCAGATTICA TTTGTAAAAT TGATTGTAGA	11520
CTACATTGTA GAAATGGTT TTTTAGAGAC GAAAGTTTA ACGCAAGAGC CGTTTAAATC	11580
TTATGGTTCT GTTCAACTAC TCTTCCAACA CCAACTACCA GTACTTCGTA ATATTGTTCA	11640
AATCATTGAA CTTATCAATA ATCGAGCTGG AGAAGCGGCT TAAATCTTAA AGTGATTGCC	11700
ATGCTGAGAC TCATTTAAAA TTA AAAAGAG TAGAAATTTA TGCTATATAT GAGAAGTTTT	11760

462

ATTAGGAAGA ATGTCATCGT TTTCTAGAA TACAGTATCA GTTGTTAAGT GGTGATAAA	11820
TTTCAAAGTA GATACTTGTA CCACGATGTT TGTGATCGA GTTATTAAACA AAAGAGCTAC	11880
TTTGATTTTA AAGAAATAGA AAACAAAAAG CCGAGCAAGA ATTCAATTGC AGGAGAAAAT	11940
GAAATAATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAA CTAGCTGCAG GCTGCTCAAA	12000
ACACTGTTTT GAGGTTGCAG ATGGAAGCTG ACGCGGATTG AAGAGATTTT CGAAGAGTAT	12060
AAATCTTCCT AGGATAAAGC AAAACGCATA GTATCAAGGG TTTTCAACAC TTGATACTAT	12120
GCGTTTTCTG ATGTTAAAGA CTTTCTACCA GGTTTTTTAA AAGCATAATT GTTAGTTGTA	12180
GTCAATTTATT ATTCTTCAA GAAAAATGGT GGGGCGAATT TTTTCAGTTC TTCAAAGCAC	12240
TTTGTAGCAG TATCTGCATC TTCACAGATG ATAAGACAGA CATCATTACC ACAAAGGGTA	12300
GCGATAGCGT CAGGGAAGCT CAAAGTATCA ATGATAGAAC CAAAGGATTG AGCCAGTCCA	12360
GGAAGGGTTT TTAGTAGGAC TTGGTGTGA ACTGGGCGCA TCCAGACAAG GGCGTCTTCC	12420
ATGTAGAGTT CGAGACGTTT TTCCCATTTT GAGATGGAAC CATTGTTAAG AACATAATAA	12480
GCGCTATCTT CTTGCGGAC TTTTGATAGG TTCATATTTT TGATGTCGCG TGAGAGGGTT	12540
GCCTGGGTTA CTTGAATGTC GTTCTCAGCA AGAAGGGCTT GCAACTCAGC CTGTGTATGA	12600
ATCTTGTTTT TTGTGATAAG AGCGCGTATA AGTTGGTGGC GGTGTTCTGA TTTATTCTA	12660
ATAATGTAAC TCCTTTTAGC AAGGTAAGGT AAGCATGGAC TGAGCGAGGT CGACAGTCAA	12720
GTGGTAGTCT GTATTGTCAC GGATGGTGAT TTCAAAGTCA GTAGTATAGA GGAATAAACG	12780
GAGAGTGTCT CCTTCTTTTA GCTTGTAAT AGTTGGCTGC AGTTCAAATT GAACGTCCAT	12840
CCATTCACTT GCAGTAATAT CCTCTACTAA CAGTAAATCA TTTCTATTTT GTAAATTAAG	12900
GTAACCTTTT GTCACGACTC GTTGTGCCTC TGGTCTAAAT GGCAATTCAC AGAGATTTTC	12960
CAACATGTGA TAGCGACCGT TGTCAATGGT TCTAGCACTT AAAATAGCTG GATAAGGTTG	13020
TAGGTATTTT TTTTGCCCAA ATTCTAGCAG TTGGGCAGAT AAGAGCCCCT TGTTTGTACT	13080
GGATTTGATA CGAAGATTGA GCTGAGCGCG ACCGTTTAGG TGGAGATCTT TAGTCACAGG	13140
AAGGTTAATA GTAATCTGAT TGGCTTTCCC TTGATAGAGC TCTGTATTGA AGGTTTGGTA	13200
TGTCTTACCA TAGCGCTCAA AATCCTTATC TGGGTACTGG TTTTGAATAG CTTGCTCTTC	13260
TTGACCAAGT GAGAAGGTTT CACAGTTTTT TTGCCCACCG AAGTTATCAA GTGATAACCA	13320
AGTCTGTGGA GCTGTATTGT CCTGCCAGAT AACAGTAGGA AGTTGAAAGT CTGTTTCCTG	13380
TCCTAGTAAT TTCTTGGTCA ATAAGGCATT TATGGACTCA CGGAAGTCAA TTGATTGCCA	13440
ATTGTTTATG TAAACATGGG CACCATTATG GAAAAAGAGA TGCTTGTGTA TATGAGTAGG	13500
AAGAGCATGG AACATCTGGT AAACATGAAG TGGTTTGACA TTCCAATCCT GAGAACCATG	13560

463

AGTAAAGACA ACCTCTGCCT TTACTTPTATG GGCATTGAGC AGATAATTGC GGTCAATGCCA	13620
AAACTGATTG TAGTCCCCAG TTTTTCGGTC TAGCTGAGCT TTCACTTTTT CTAAGTCAGC	13680
TTGGTGAGCT TCATTGCCAC GGATATAGTC GCCAGCTAAG AGATTACGAG AATAGGTTAA	13740
CTCAGCAAGG GAGTCAAAGT CCTCACCTGG ATAACCACCT GGGCTAGTCA CCAGACCGTT	13800
TTACCGGTAG TAGTTGTACC ATGATGAAAT TCCTGCCTCG GCAATGATAA CTTCTAAACC	13860
ATCGACTCCT GTAGTCGCAA GACCATTGGA CATGGTACCT AGATAGGAAA GTCCTGTTGT	13920
AGCAACTTTT CCGTTTGACC AATCAGCCTT GACTTGACGC TGGCGCGTGT GATCAGTAA	13980
GGCACGGCAA CGACCGTTAA GCCAATCGAT GACATTTTTA TAAGCCTCGA TTTGCTGGTA	14040
GTCTCCATTA GTCATGAAAC CTGTCGAGTC TTTGGTACCA ACACCTGAGA CATAGAGATT	14100
GGCAAAGCCT CTCGGAAGGA AGTAGTCGTT TAGTGTATAG CTAGAGTTGA TGTGAGTTAG	14160
CTTTTCCTCA GCCTCTGCTA TAAGCTCAGC TTTACCTTGG GGTGCGACGA GATTTAGTTG	14220
AGGTTTCTCT AGCTCAATCT TGTGAGGAAG CTTAACCTCA AGCTCGCCCT CCATCTTGTA	14280
GAGAGCCTTG TCACTAGCCT TGTCAATTGGT TCCCTGATGA TAAGGGCTGG CTGTCATGAT	14340
GGCAGGGATT TTTCCATCAA AACGAGGCG AATAATGCTA ACCTTTACTA GGTCTGATAG	14400
CCCTTTTGG TCAGTATCGA CACGAGACTC AACGTAAACG ACTTCACGAA TGACATCCTG	14460
GTTAGAAAAA GTAGCCAAAC TCTTGCCGTT AAAGTAGTGG TAGTCATTAT CCTCCGGAAT	14520
AAGACCATCA CTAACAAGTT GGTGATAAG AGTATTTCTT TTTTGGTGC GAGTATTGAG	14580
TAACGTATAG AGATTTTCAA TCAAGTCACC ATATATAATG GGAAATCCAG TTTCTTTACG	14640
AAAAACGTCA CTATCTTCGA AGTCAACCAA ATAAGAAAAG CCTAAAAGTT GAAAAGCAAC	14700
AGTATAAAAA ATATCTGCTG TCAGTTCATC TTCTGATTGA AAAAATGTCA GCAGGTCTGT	14760
TTTTTTATCA GCTGCTAGGA TAGAAAGTGG GTAGTTGGTG TCTTGATAAG TGAAAAAGAA	14820
ACGACGTAAA AAGGTTTCAA GTGAGTCTTT GTGATTGGCT GTATTTTGTA AATCAAAGCC	14880
ACATTTTTTT AGTTCAGATA AGACATTTTC TTTTGAAAA TTGATATAAC TATATTGATT	14940
AAAACGCATA GAACCTCCAT ATAGAATGAC AGTTAAGGTT ATTATATCAA AAAAAAGCA	15000
GAAAGGGAAT TGTAACTTC AAAAGGAAAT AATCCAATAA AAATGAATAA AGTACTAAAT	15060
TCAATATAGA GAACAGAGTA ACAATAAGAA TAAATAGATA GGGTATAAAA GTTCTAGGAG	15120
ATTTATATTA TATGCTTTCT ATTTTATAT ACAATATAGT ATAAATATAA AAATGATGAC	15180
AAAAATACAA ATGAATAGAA AATAAATTAG TAAGCTGATG AAATTTTCT CAAGAGAAGC	15240
CATTTATAGG TGAAATGGT ATAATATAGT GAGAAGGATA GAGGAGAAGT GTAAATTGAT	15300

464

CGCACAACTA GATACAAAAA CAGTCTATAG TTTTATGGAA AGCGTCATTT CGATCGAAAA	15360
GTATGTGAGA GCAGCTAAAG AATACGGCTA CACTCATTTG GCTATGATGG ATATTGACAA	15420
TCTTTATGGC GCTTTCGACT TTCTAGAGAT TACAAAAAAA TACGGCATTG ATCCTTTGCT	15480
AGGGCTTGAA ATGACAGTGT TTGTAGATGA TCAGGGAGTG AATTGCGCT TTTTAGCTCT	15540
ATCTAGTGTG GGCTATCAGC AGTTGATGAA GCTTTCGACA GCCAAGATGC AGGGGGAGAA	15600
AACTTGGTCA GTCCTGTCCC AGTACCTGGA GGATATCGCG GTCATTGTGC CTTATTTTGA	15660
TAGAGTTGAG TCGTTAGAAC TAGGCTGTGA TTACTATATA GGGGTTTATC CAGAAACACT	15720
AGCAAGCGAA TTTCATCATC CTATCTTACC TCTTTATCGG GTCAACGCTT TTGAAAGCAG	15780
GGATAGAGAA GTTCTTCAAG TTTTAACAGC GATTAAAGAA AATCTACCGC TCAGAGAAGT	15840
TCCCTTGCGT TCGAGACAAG ATGTCTTTAT ATCAGCAAGT TCTTTAGAGA AACTATTCCA	15900
AGAGCGTTTT CCGCAAGCTT TGGACAATTT AGAAAAGCTT ATTTTCAGGCA TTTCTTACGA	15960
CTTGGATACT AGTCTGAAAC TGCCTCGTTT TAATCCAGCT AGACCAGCAG TAGAGGAGTT	16020
GAGAGAGCGT GCTGAAC TGG	16080
GAGAGAGCGT GCTGAAC TGG	16080
TAGACTAGAC CAAGAATTGT CTGTTATTCA TGATATGGGC TTTGATGATT ATTTCTTGGT	16140
TGTTTGGGAT TTGTTGCGTT TTGGACAATC GAATGGCTAT TATATGGGAA TGGGAAGGGG	16200
TTCTGCAGTA GGCAGTTTGG TTTCTTATGC CTTAGACATC ACGGGGATTG ACCCAGTAGA	16260
GAAAAATCTG ATTTTGAAC GCTTTCTTAA TCGTGAACGC TATACCATGC CTGATATTGA	16320
TATTGATATC CCAGATATTT ATCGTCCAGA TTTTATCAGA TATGTTGGTA ATAAATATGG	16380
TAGTAAACAT GCGGCACAAA TCGTTACTTT TTCAACCTTT GGAGCCAAGC AAGCTCTTCG	16440
AGATGTCTTG AAACGCTTTG GTGTGCCAGA GTATGAATTA TCTGCAATTA CTAAGAAAAT	16500
CAGTTTTCGT GACAATCTTA AGTCGGCCTA TGAGGGAAAT CTCCAGTTTC GTCAGCAAAT	16560
CAATAGTAAG TTAGAATACC AAAAAGCTTT TGAGATTGCT TGCAAGATAG AGGGCTATCC	16620
AAGGCAAACC TCTGTCCATG CGGCTGGTGT TGTAATTAGT GACCAAGATT TAACCAACTA	16680
CATTCTCTA AAGTATGGTG ATGAAATTC ACTGACTCAG TATGATGCTC ATGGAGTTGA	16740
GGCTAGCGGA CTTTGAAGA TGGACTTCT GGGACTACGA AATTGACCT TTGTCCAGAA	16800
GATGCAAGAG TTGCTTGCTG AAACAGAAGG TATTCATCTG AAAATTGAAG AAATCGATTT	16860
AGAAGACAAA GAAACGTTAG CTTTATTTGC CTCTGGTAAT ACAAAGGTA TCTTTCAATT	16920
TGAGCAACCA GGTGCCATTC GTCTGCTTAA GCGTGTGCAA CCAGTCTGTT TTGAAGATGT	16980
CGTCGCGACT ACTTCTCTAA ATCGACCGGG TGCTAGTGAC TATATCAATA ATTTTGTGGC	17040
AAGAAAGCAT GGGCAGGAAG AAGTGACTGT TCTGGATCCA GTACTGGAGG ATATTTTGGC	17100

465

TCCAACCTAC GGCATAATGC TCTATCAGGA GCAGGTTATG CAGGTTGCCC AGCGACTTGC	17160
CGGATTTAGT CTGGGAAAG CCGATATTTT GCGTCGGGCT ATGGGGAAAA AGGATGCCTC	17220
TGCCATGCAT GAGATGAGGG CTTCTTTTAT TCAAGGTTCA TTAGAAGCTG GTCATACTGT	17280
GGAAAAAGCA GAGCAGGTCT TTGATGTTAT GGAGAAGTTT GCAGGTTATG GTTTTAACAG	17340
GTCACACGCC TATGCCTACT CAGCCTTGGC CTTCCAGTTG GCTTATTTCA AAACGCATTA	17400
TCCAGCCATT TTTTATCAGG TCATGTTAAA TTCTTCCAAC AGTGATTACT TAATAGATGC	17460
ACTTGAAGCA GGTTTTGAAG TAGCCTCTCT ATCCATCAAC ACCATTCCCT ATCACGATAA	17520
AATTGCCAAC AAGGCCATCT ATCTAGGTTT GAAATCCATT AAAGGAGTCA GTAATGATTT	17580
AGCTCTCTGG ATTATTGAAA ATAGACCTTA TTCTAACATT GAAGATTTTA TAGCTAAATT	17640
ACCTGAGAAT TATCTGAAAC TTCCTCTGCT AGAACCTTTG GTAAAAGTTG GTCTTTTTCGA	17700
TTCATTTGAA AAAAATCGTC AAAAAGTATT TAATAACTTA GCTAATCTAT TTGAATTTGT	17760
GAAAGAGTTG GGAAGTTTGT TTGGAGATGC TATTTATAGT TGGCAGGAAT CGGAAGATTC	17820
GACGGAACAA GAAAAATTTT ATATGGAACA AGAGCTTTTA GGGATAGGTG TCAGCAAACA	17880
TCCACTACAA GCTATGCAA GTAAGGCTAT TTACCCGATT ACCCCAATCG GAAATTTGTC	17940
AGAAAATAGC TATGCTATTA TCTTGSTTGA AGTTCAGAAA ATAAAAGTGA TTCGTACCAA	18000
AAAGGTGAA AATATGGCCT TCTTACAGGC AGATGATAGT AAGAAAAAAT TGGATGTCAC	18060
TCTCTTTTCA GACTTATATC GTCAGGTGG ACAGGAAATA AAAGAGGGAG CCTTCTACTA	18120
TGTAAAAGGA AAAATACAAT CACGTGATGG CCGTCTGCAA ATGATTGCAC AAGAAATAAG	18180
AGAAGCAGTT GCTGAACGCT TTTGGATACA GGTGAAAAAT CATGAATCGG ATCAAGAAAT	18240
TTCACGCATT TTAGAACAAT TTAAAGGCCC AATCCAGTC ATCATCCGGT ATGAAGAGGA	18300
ACAGAAAACC ATCGTTTCTC CCCATCATTT TGTAGCTAAA TCCAATGAAT TAGAGGAGAA	18360
ATTGAATGAA ATCGTTATGA AAACGATTTA TCGCTAAAAA TACGGAAAAAT AGAAGAATTT	18420
TCAACGTAAA TGTGGTATAA TCAGTAAGAA TGTTAAAAGA AAAAGGAGCA TAACCAATAT	18480
GAAACGTATT GCTGTTTGA CTAGTGGTGG AGACGCCCCCT GGTATGAACG CTGCCATCCG	18540
TGCAGTTGTT CGTCAAGCAA TTTCAGAAGG AATGGAAGTT TTTGGTATCT ATGACGGATA	18600
TGCTGGTATG GTTGCCGGTG AAATTCATCC CCTAGATGCA GCTTCAGTAG GGGACATCAT	18660
TTCTCGTGGT GGTACTTTCC TTCACTCAGC TCGTTACCCA GAGTTGCTC AACTTGAAGG	18720
GCAACTTAAA GGGATTGAGC AATTGAAAAA ACACGGAATT GAAGGTGTAG TTGTTATCGG	18780
TGGTGACGGA TCTTACCACG GCGCTATGCG TTGACTGAA CATGGCTTCC CAGCTATTGG	18840

466

TCTTCCAGGT ACAATCGATA ACGATATCGT TGGTACTGAC TTTACAATCG GTTTTGACAC	18900
AGCGGTACT ACTGCCATGG ACGCTATCGA TAAGATTCGT GATACATCAT CAAGTCACCG	18960
TCGTACTTTT GTAATCGAAG TTATGGGACG TAACGCTGGT GATATCGCTC TTTGGGCTGG	19020
TATTGCAACT GGTGCTGATG AAATCATCAT CCCTGAAGCA GGCTTCAAGA TGGAAGATAT	19080
CGTAGCAAGC ATCAAAGCTG GTTATGAATG TGGTAAAAA CACAATATTA TCGTCTTAGC	19140
TGAAGGTGTG ATGTCAGCGG CTGAATTTGG TCAAAAACTT AAAGAAGCTG GAGATACAAG	19200
CGACCTTCGT GTAACAGAAC TTGGACATAT TCAACGTGGT GGTTCCTCAA CTGCGCGTGA	19260
CCGTGTTTGG GCGTCACGTA TGGGTGCACA TGCTGTTAAA CTTCTTAAAG AAGGTATCGG	19320
TGGTGTGCG GTTGGTATTC GTAACGAAAA AATGGTTGAA AATCCAATTC TTGGTACTGC	19380
AGAAGAAGGG GCATTGTTTA GCCTTACTGC AGAAGGTAAG ATTGTGGTTA ACAACCCAGC	19440
TACAAA	19446

(2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16593 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TCGTAAATAT GCTCTGTTTT TGGATTTTGT TTCTTAATCT GTTTGGCAAG TGCCTTCATC	60
ATAGAAATAG GACCACACAT ATAGACGGTT GCATGTTGCG GCACTTCTTT TTGTTCAAAA	120
TTAAGATAGC CGTCTTTCGT ACTGTCGATT AGATGGAGTT CAAAATTAGG ATTTTCTGA	180
GCATAGTTAC GGAGTAAATC TAGGTAGACT GCATTTTCAT CTCCACGGAA GCTATAGTAG	240
AAGTGAACCT GTTTATCTAA AATAGGATGT TCACGGATGT AAGAGATGAA GGGGGTGATC	300
CCAATACCTC CAGCAATCCA AACCTGATTT TCTCGTCCTT CTTCTATGAT CATGTGTCCG	360
TAAGCTCTGT CTAGGGTTAC TTTGCTGCCC GCTTGAAGAT TATCATAGAT ATTCTTG GTA	420
TGGTCGCCTG AAGTTTAAAC AGTAAAGTAA AGAGTTTGAC CATGACCTCC TGAGATAGAA	480
AAGGGATGCG GAGCACTTTC AAAGCCTTCT TGGAAAATCT TTAGAAAGGC AAATTGTCCT	540
GATTGATAGT TGAAAGGTCT GCTAAGATGG ATTTGAATTT CTCTAGTATC GTGATTAAAG	600
CGTTTGAGAT GGGTAATTTT CCCTAGATAG GGGAAAGGAA TCTTTTGATA TAGAAAAATG	660
ATATAAAAC CAGCTAGTAA GCCTAAAAGG GCATAGCTAC CAACAAGAAA ACTTAGAAGA	720
TTAAATGTAA GGAGACGATT GCCCATATC ATGTAGATGT GAAAGAGTCC TAAATATAG	780

467

GCTAGGTAAA	CCAGGCGGTG	AATCCATCGC	CAAGCTTCGT	ATTGGATGTA	TTTGCCTAAA	840
TAGGCGACAA	GGATGATGCT	GGCAAAGATA	TAGATGGCAA	GATTGCCAAA	CTGAGCAGCT	900
AAGCGAGAGC	CCCACAAACC	GCCCATACTA	AAGTTATGAA	AGATTAGTAG	GATGATTGAG	960
AGAAAGGCTG	TGAATTTGTG	GACGGTGTAG	ACCTTCTCCA	AACTGTGAAA	CCAGCTTTCT	1020
AGTAGTGGGA	GACGAGTGGC	TAGGATAAAA	GTCAGAGATA	GGCTTGTTAA	AGCTAGTCCT	1080
GGAATCATGA	ATTGGGGAGA	AGTGTTTCATC	CAAGTCAAAA	GAGTCAAGAT	AAAAC TAGCT	1140
ATGATAAAGA	GTAGTCCTTT	GA CTGATTTC	ATAGAAAATT	CCATTTTCATT	TAGATTTCGA	1200
TTTGTGTGTA	ATAAATTTGT	TACATTTTAT	CATAGAAAAT	GTATGGTGTC	AAATTGAGGT	1260
CTATAAATAT	CTACTCTCAT	CAAAAACTC	TCCAATTGAA	CTGGAGAGTG	GCTGTTTATA	1320
CTCAATGAAA	ATCAAAGAGC	AAACTAGGAA	GCTAGCCGCA	AGTTGCTCAA	AACACTGTTT	1380
TGAGGTTGCA	GATAGAGCTG	ACGTGGTTTG	AAGAGATTTT	CGAAGAGTGT	TATTCTGCAG	1440
CTTGTTGCCA	ACGTTTGGCT	AGCATATGAG	ACAGGCTAGA	AATTGCTAGG	TTAAAGCTGA	1500
AGTAGATGAG	GGCAATCAGG	ATGTAAAGAC	TGAAGACCTG	CTCTGGTTTCG	AAATAACGGC	1560
CCATGAGAAT	TTGGCTGGCT	CCAAAGAGTT	CTTGTAGGGC	GATAACAGAG	TAGAGGAGAC	1620
TGGTATCCTT	AATCACGGTA	ACAAACTGAG	AAATGATGGC	TGGTAGCATT	TTGCGGATGG	1680
CTTGTTGGAG	AATGATGTAG	TAGAGGATTT	GGGCTGAGGT	GAAGCCTTGT	GACATTCCTG	1740
CTTCGTA CTG	TCCCTTGTCT	ACGGCATTGA	GACCGCCTCG	AATAATCTCA	GCCAAGGCTG	1800
CTGATGTA AA	GAGAGTAAAG	GCTGTAATAC	CTGCTGGTGT	GGATTTTCATT	TTGAACACCA	1860
AAAAGATAGT	AAAAATCCAG	AGAAGGTTGG	GAACGTTGCG	CACAAACTCG	ATATAAATAC	1920
TGGAAATAAT	GCGTAAGACA	GGATTTTTCG	CATTTCTCGT	GACAGCTAGC	ACCGTACCGA	1980
TGATAGTAGA	GAGGATGATG	GCAATCAGAG	AAATATAGAG	GGTCAAGCCA	AATCCTTTAA	2040
AGATAAAGAC	TAGGTTATCT	GGGGTTAAAA	CTTCTAAAAT	AGATTCCATA	GTAACCTCCT	2100
AAAGTGAATA	GGCTTTTTCG	TTGGCTTGCT	CCATCTTGCG	ACCAA ACTGG	GCAACAGGGA	2160
AGCATAGAGC	AAAGTAGAGA	AGAGCAGCAC	CTAAAAAGGC	TGGTATATAG	TTTCCGTTGA	2220
GAGCCGACCA	AGACTTAGTC	ACAAACATCA	AGTCTACTCC	AGAGATGATA	GCTACAGTAG	2280
AGGTGTCTTT	GATGAGGTTA	ACAATTTGGT	TGGTCAATGG	AGGGAGAATG	ATGCGGAAGG	2340
CCTGAGGCAA	GATAATCAAG	CGCATGGCAC	TGATATAGGT	AAAACCTTGC	GACAAGGCGG	2400
CCTCCATCTG	ACCACTAGGA	ATAGACTGAA	TCCCTGAACG	AATAACCTCA	GCGATATAAG	2460
CGCCGTGATA	GAGTCCCACG	CAGAGAACGG	CTGTCCAATA	AATTGGAATC	ATGATGATAT	2520

468

GGTCACTGAT AAGAGGTAGG CCATAAAAA CAATAACAAA CTGCACCAAG AGGGGAGTAT	2580
TTTGGTAAAA TTCAACAAAG ATGCGAGCTA AAATGCGTAA AATTGGACGT TTACTGGTTG	2640
ACATGGCACC AAAGAAGATG CCCAAAACCA TAGCGAGGAT AAAGGAACCA ACCGCTAGGG	2700
CAAGGGTGAA GAGGAAACCA TTGAAAAATT GTCCAAAATC CTGAAAATAG GCTGTCCAAG	2760
ATGATAAATC TGTCATGGGG TGTCTCCTT AATCTGCAGT ATGGCTAGAT GGTTCGAGCT	2820
TGTAACGGTC ATAAAGTTTC TGCAAACTAC CATCCTTGCT CCAATTAGTA ACCAAGTTAT	2880
CAAGATAGTC GTTGAGCTCT GTATTTGATT TCTTGGTAAC AATACCGTAG TCAGATGGCT	2940
TGAAACTATC ATCTAGTAGT GCTGTCCGTT TACTAGTGA GCCAGATAGA ATAGAGCGGT	3000
CAACGGAAAA GGTATCGATA CGATGAGCGT GCAGGGAAGT AATCAATTCT GGGTAGGAAC	3060
CAAGTTCGAC GAATTTAAAC TTCAGACCTT TCTTTTTACC CAGTTCAGTA ATCAGGCGTT	3120
GGGTGATAGA ACCTTGGGCG ACTCCGATGG TTTTGCCGTT TAGGTCTCA ATCTTTTTGA	3180
TTTTGGCAGA TTTATTGACC AAAAATCCAG AAGCGTCTGT GTAGTAGGGA CTGGTAAAGT	3240
TGTAGAGTTT TTTGCGTTTCG TCCGTGATGG TAAAGGTCGC GATATCCATA TCGACCTGTT	3300
CATTGTCTAG AAGGGGGCCG CGGGTTTGTG CTGTAACCGG CACATAGCGA ATCTTGACCT	3360
TGAGTTCATC AGCTACCATC TTGGCCAAGT CGGTTTCGAT ACCAGAATAA GTACCGGTCT	3420
TGGGATCTTT GTAACCAAAA TTGGGAACGT CTGTTTTGAC ACCGACAACC AGTTCGCCTC	3480
TTTTTTGAAT GTCTGCGATA CTTGTATCAG CCTGGACTGG TTTGGCAGCA GCAAGGCCGA	3540
AAAGGCTAAT CAATAATGCT GATAAAAAGA ATTTTTTTTC ATAGCGCCTT CCTTATTGA	3600
CTTTGTCACT TTCGTGGTTG ATAATTTTGC TGAGGAATTG TTGGGCACGA GGTTCGCTTG	3660
GATTGTCAAA AAAGTTATCG ACATCTGTCG TATCTACTAA AACTTCTCCG TCGGCCATAA	3720
AGATAATGCG GTCCGCAACC TCTCGAGCAA AGCCCATTTT GTGGGTAACG ATGATCATGT	3780
TCATCCCATC ATGCGCCAGT TTCTGCATAA CTGCTAGAAC ATCTCCGATA GTCTCAGGAT	3840
CAAGAGCAGA TGTGGTTTCA TCAAAGAGGA GGAGTTCCGG ATGCATAGCA AGACCACGAG	3900
CGATGGCGAT CCGCTGTTTT TGTCACCAG ATAGCATGGC GGGATAGGAA TCTTTCTTGT	3960
CCCACATATT TACAAATTCC AGATATTTTT GGGCGGTTTT TTCAGCTTCT TTTTATCAA	4020
TTCTTAGAAC TTCAATGGGT GCAAGCGTTA CGTTTCTTAA CACAGCTTTG TGTGGATAAA	4080
GGTTAAAAATG TTGAAAAACC ATGCCGACTT CCTTGCGAAG AGGTACCAAA TCTTCTGGC	4140
TGGCACCAGC AACTTGGTGC CCATTGACTA GGAGACTTCC TTTGTCAACA GTCTCTAAAC	4200
CATTGATCGT ACGGATAAGA GTGGACTTCC CAGAGCCAGA AGGTCCAAGC AGGACAACAA	4260
CTTGTCTTTT TTCAAAACGG AGATTGATGT TCGGAATGC GTGGTAGTCT CCGTAATATT	4320

TTTCGACGTT	TTTAAATPCT	ACTAAAGCCA	TGAGAGATCT	CTATTGTGTT	ATATTTTATA	4380
ACACGGTTCT	ACAATAAAG	AATGTTCTTG	TCAAATCATA	TCTGAAAAA	TTCACATATAG	4440
TGAAATAAGA	ACAGGAAAA	TCGATCGGGA	CAGTCAAATC	GATTTCTAAC	AATATTTTAG	4500
AAGTAGAGGT	GTAATATCT	AGTTTCAATA	TACTATAAAA	TGTTATAAAA	AAGCAATCTG	4560
GATAGAGAAA	ACGTCTAAAT	CATGTTATAA	TGAAGCAATA	GAATTCCTAG	AAAGAGTGGA	4620
TGTCCTTTTG	ATAACACCTA	CTTATGAATG	GCAGTTTGCC	CTGCAGGTAG	AAGATGCGGA	4680
TTTTACAAAG	ATAGCCAAGA	AGGCTGGACT	GGTCTCTGAG	GTGGCTCGGT	TATTGTTTGA	4740
GAGAGGGATT	CAGAACCAAG	AAAGTCTGAA	GAAGTTTTTA	GAACCTTCCT	TGGAGGACTT	4800
ACATGATGCT	TATCTGCTCC	ATGATATGGA	CAAGGCAGTG	GAGCGGATTC	GTCAGGCTAT	4860
TGAAGAAGGG	GAAAATATTC	TTGTTTATGG	AGACTATGAT	GCGGATGGCA	TGACTTCGGC	4920
TTCTATGTG	AAGGAAAGTT	TGGAACAAC	TGGTGCTGAG	TGCCGAGTTT	ACCTGCCAAA	4980
TCGTTTACC	GATGGCTATG	GCCCTAATGC	TAGTGTTTAT	AAATACTTTA	TCGAGCAAGA	5040
AGGGATTTC	TTGATTGTGA	CGGTGGACAA	TGGGTTGCT	GGTCATGAGG	CTATTGCATT	5100
GGCTCAGTCT	ATGGGAGTAG	ATGTCATTGT	GACAGACCAT	CATTCCATGC	CTGAAACCCT	5160
GCCAGATGCT	TATGCTATTG	TCCATCCTGA	ACATCCAGAT	GCGGATTATC	CTTTTAAATA	5220
TTTGGCTGGT	TGTGGAGTTG	CTTCAAGTT	GGCTTGCGC	CTGTTAGAAG	AAGTGCAAGT	5280
GGAATTGCTT	GATTTGGTCG	CTATTGGAAC	TATTGCAGAT	ATGGTGAGTC	TGACGGATGA	5340
AAATCGTATC	TTAGTTCAAT	ATGGTCTGGA	AATGTTGGGT	CATACCCAGC	GCATTGGTCT	5400
GCAAGAAATG	CTGGACATGG	CTGGGATTGC	TGCCAACGAA	GTAACAGAAG	AAACGGTTGG	5460
TTTCCAGATT	GCTCCTCGTT	TGAATGCCTT	GGGTCGCTTG	GATGATCCCA	ATCCTGCCAT	5520
TGATTTGTTG	ACTGGATTG	ATGATGAGGA	AGCGCATGAG	ATTGCCCTTA	TGATTACCA	5580
GAAAAACGAA	GAGCGCAAGG	AAATCGTTCA	GTCTATCTAT	GAAGAAGCCA	AGACCATCGT	5640
GGATCCTGAG	AAGAAGGTTT	AGGTCTTGGC	CAAGGAAGGC	TGGAATCCTG	GGGTTCCTAGG	5700
AATCGTGGCT	GGTCGTTTAT	TGGAAGAATT	GGGACAGACA	GTCATTGTTC	TTAATATAGA	5760
AGACGGTCGT	GCCAAGGGCA	GTGCTCGTAG	TGTGGAAGCG	GTCGATATTT	TTGAAGCTCT	5820
GGATCCCCAT	CGAGACCTCT	TCATCGCCTT	TGGAGGTCAT	GCAGGTGCAG	CGGGTATGAC	5880
GCTGGAAGTT	GAGCAACTCT	CAGATTTATC	TCAGGTTTTG	GAAGATTATG	TTCGTGAAAA	5940
AGGTGCAGAT	GCTGGTGGCA	AGAATAAGTT	AAACCTAGAT	GAAGAGTTGG	ATTTGGAGGC	6000
ACTTAGCTTG	GAAACGGTCA	AAAGTTTGA	ACGTTTAGCT	CCTTTTGGAA	TGGATAATCA	6060

470

GAAACCTATT	TTTTATATCA	AGAATTTTCA	GGTCGAAAGT	GCTCGTACTA	TGGGGGCAGG	6120
TAATGCCCAT	CTAAAGCTGA	AAATTTCCAA	GGGTGAGGCG	AGTTTTGAAG	TGGTAGCCTT	6180
TGGTCAAGGC	AGATGGGCGA	CAGAGTTTTC	TCAAACCAAG	AATCTAGAGT	TAGCGGTTAA	6240
ATTGTCTGTC	AACCAATGGA	ATGGCCAAAC	TGCCCTCCAG	TTGATGATGG	TGGATGCGCG	6300
AGTGAAGGT	GTTCAACTTT	TTAACATTCC	TGGAATAAAT	GCAGTCTTGC	CAGAAGGTGT	6360
TCCAGTCTTG	GATTTTCCTG	GAGAAGTCCC	AAATCTTGCG	GCTAGTGAAG	CTGTGTGCTG	6420
AAAAAACATT	CCAGAGGATA	TTACTCAGCT	GAAGACCATT	TTTCAGGAAC	AGCATTCTTC	6480
TGCTGTCTAT	TTCAAAAATG	ATATTGACAA	GGCTTATTAT	CTGACAGGTT	ATGGGACTAG	6540
AGATCAGTTT	GCCAAATTGT	ACAAGACTAT	TTACCAGTTC	CCAGAGTTTG	ATATTCGCTA	6600
CAAGCTGAAA	GATTTGGCTG	CATATCTTAA	TATTCAACAA	ATCTTGCTGG	TCAAGATGAT	6660
TCAAGTATTT	GAAGAACTAG	GCTTTGTGAC	GATAAAAGAT	GGTGTGATGA	CAGTCAATAA	6720
AGAGGCGCCA	AAGCGGGAGA	TAGGAGAAAG	TCAAATTTAC	CAAAATCTCA	AACAAACCGT	6780
TAAAGACCAA	GAAATGATGG	CGCTGGGTAC	GGTGCAAGAA	ATTTATGATT	TTTGATGGA	6840
AAAAGAGTAG	AAGTTAGGAA	AGAGTTGGGA	AATCAACTCT	TTTTTGAAAA	CAGACCTTCA	6900
TTTTGAAAAT	CATCAAAAAA	ATGGTATAAT	GGTAGGAAAA	GATTCGGCTG	AAAGTATCAG	6960
AACTTTTAGA	ATAAGAGGGT	AGAATTGCCC	TATAATCAAG	ATAAACTAAG	ATTTTGGAGG	7020
AAAAATGAGT	AATATCAGTT	TAACAACACT	TGGTGGTGTG	CGTGAGAATG	GAAAAAATAT	7080
GTACATTGCT	GAAATTGGAG	AGTCCATTTT	TGTTTTGAAT	GTAGGGTTAA	AATATCCTGA	7140
AAATGAACAA	TTAGGGGTCG	ATGTGGTGAT	TCCAAACATG	GATTACCTTT	TTGAAAATAG	7200
CGACCGTATT	GCTGGGGTTT	TCCTGACCCA	CGGGCATGCG	GATGCCATTG	GTGCTCTACC	7260
GTATCTCTTG	GCAGAGGCTA	AAGTTCCTGT	ATTTGGGTCT	GAGTTGACCA	TTGAGTTGGC	7320
AAAGCTCTTT	GTCAAAGGAA	ATGATGCCGT	TAAGAAATTT	AATGATTTC	ATGTCAATGA	7380
TGAGAATACG	GAGATTGATT	TTGGTGGGAC	AGTGGTTTCC	TTCTTCCCTA	CGACTTACTC	7440
CGTTCCAGAG	AGTCTGGGAA	TTGTCTTGAA	GACATCGGAA	GGAAGCATCG	TTTATACAGG	7500
TGACTTCAAA	TTTGACCAAA	CGGCTAGTGA	ATCTTATGCA	ACTGATTTTG	CTCGTTTGGC	7560
AGAGATTGGT	CGTGACGGCG	TCCTGGCTCT	CCTCAGTGAT	TCGGCCAATG	CAGACAGCAA	7620
TATTCAGGTG	GCTAGTGAAA	GTGAAGTTAG	GGATGAAATT	ACCCAAACTA	TTGCTGACTG	7680
GGAAGGTCGT	ATCATCGTTG	CAGCTGTTTC	CAGTAATCTT	TCTCGTATTC	AGCAGATTTT	7740
TGACGCTGCG	GATAAAACAG	GTCGACGTAT	CGTCTTGACA	GGATTTGATA	TTGAAAATAT	7800
CGTCCGCACA	GCGATTGCTC	TTAAGAAGTT	GTCTTTAGCC	AACGAAATTC	TTTTGATTAA	7860

471

GCCTAAAGAT ATGTCTCGCT TTGAAGACCA TGAGTTGATT ATTCTTGAGA CAGGTCGTAT	7920
GGGTGAGCCT ATCAATGGAC TTCGTAAGAT GTCGATTGGT CGCCATCGTT ATGTAGAAAT	7980
CAAGGATGGG GACCTAGTCT ATATTGCTAC GGCTCCGTCT ATTGCTAAAG AAGCCTTTGT	8040
TGCGCGTGTG GAAAATATGA TTTATCAGGC AGGTGGGGTT GTCAAATTGA TTACCCAAAG	8100
TTTACATGTA TCAGGGCAGC GAAATGTGCG TGATTTCAG CTGATGATCA ATCTTTTGCA	8160
ACCTAAGTAC CTCTTCCCTG TCCAAGGGGA GTATCGTGAG TTGGATGCTC ACGCTAAGGC	8220
TGCCATGGCA GTTGGGATGT TGCCAGAACG CATCTTCATT CCTAAAAAGG GGACGACCAT	8280
GGCTTACGAG AATGGAGACT TTGTTCCAGC TGGATCGGTT TCAGCAGGAG ATATCTTGAT	8340
TGATGGGAAT GCCATTGGTG ATGTTGGAAA TGTTGTTCTT CGTGACCGTA AGGTCTTGTC	8400
AGAGGATGGA ATTTTCATCG TGGCTATTAC AGTCAACCGT CGTGAGAAGA AAATTGTGGC	8460
TAGGGCTCGT GTTCACACGC GTGGATTGTG TTATCTCAAG AAGAGTCGCG ATATTCTCCG	8520
TGAAAGTTCA GAATTGATTA ACCAAACGGT AGAAGAGTAT CTTCAAGGAG ATGACTTTGA	8580
CTGGGCAGAT CTCAAAGGTA AGGTTTCGTGA CAATCTGACC AAGTACCTCT TTGATCAAAC	8640
CAAGCGTCGC CCAGCCATTT TACCAGTAGT CATGGAAGCA AAATAATCGT TGAATAAAC	8700
AGAGAGAAAG TCGAGTTTCG GCTTTTCTT ATAGAAAAAT AGAAGGAGAA AATCATGGCA	8760
GTGATGAAAA TCGAGTATTA CTCACAAGTA TTGGATATGG AGTGGGGGGT GAATGTCCTC	8820
TACCCTGATG CCAATCGAGT GGAAGAACCA GAGTGTGAAG ATATTCCCGT CTTGTACCTT	8880
TTGCACGGGA TGTCTGAAA TCATAAGT TGGCTTAAGC GGACCAATGT AGAACGCTTG	8940
CTTCGAGGAA CTAATCTCAT CGTTGTTATG CCCAATACCA GCAATGGTTG GTACACCGAT	9000
ACCCAGTATG GTTTTGACTA CTACACGGCT CTAGCAGAGG AATTGCCACA GGTTCGAAA	9060
CGCTTCTTCC CTAATATGAC GAGCAAGCGT GAAAAGACCT TTATCGCTGG TCTTTCTATG	9120
GGAGGCTACG GCTGCTTCAA ACTGGCTCTT ACGACAAATC GTTTTCTCA TGCAGCTAGT	9180
TTTTCAGGTG CCCTCAGCTT TCAAACTTT TCTCCTGAAA GTCAAAATCT GGAAGTCCA	9240
GCCTACTGGA GAGGTGTTTT TGGAGAGATT AGAGACTGGA CAACTAGTCC CTATCTCTT	9300
GAAAGTCTGG CTAAAAATC GGATAAAAAG ACCAACTTT GGGCGTGGT TGGCGAACAG	9360
GATTTCTTGT ACGAAGCCAA TAATCTCGCA GTGAAAAATC TCAAAAACT AGGTTTGTAT	9420
GTGACCTATA GCCATAGCGC TGGAACTCAC GAGTGGTACT ACTGGGAAAA ACAATTGGAA	9480
GTTTTTTTAA CAACCTACC AATTGATTTT AAATTAGAAG AGAGACTGAC TTAGTTTGAA	9540
CTTCAGCATA GGGGAGTAG AACTAAAATA AAATATGTTT TCACTAGACT TTTCAAACGm	9600

472

AAGTAGTAGA ATAGTAATAA AATACTGGAG GAAAGAGAGT AGGAAATGTA CCGTTATCAA	9660
ATTGGCATTTC CCACATTAGA ATATGATCAG TTTGTCAAAG AACATGAATT AGCCAATGTA	9720
TTACAAAGTA GTGCTTGGGA GGAAGTTAAG TCTAATTGGC AACATGAGAA GTTTGGTGTT	9780
TACAGGGAAG AAAAATTACT GGCGACAGCT AGTATTTTGA TTAGAACTCT TCCGCTAGGC	9840
TATAAAATGT TTTACATCCC AAGAGGACCT ATATTGGATT ATGGGGATAA AGAACTCTTG	9900
AATTTTGCCA TTCAGTCTAT TAAGTCCTAT GCTCGCAGTA AGAGAGCGGT TTTTGTGACT	9960
TTTGACCCAA GTATTTGCCT ATCTCAAAGT TTAATCAATC AGGAAAAGAC AGAATTTCCCT	10020
GAAATCTGG CTATTATGTA TAGTTTGCAA CAAATGGGAG TAAGGTGGTC AGGAAAAACG	10080
GAGGAAATGG GAGACACCAT TCAACCTCGT ATTCAGGCGA AAATATACAA GGAAAAATTT	10140
GAAGAAGATA AACTTTCCAA GTCAACAAAA CAGGCTATTC GAACAGCACG AAACAAAGGG	10200
CTTGAGATTTC AATATGGTGG ACTGGAAC TAAGATTTCAT TTTCCGAGTT GATGAAAAA	10260
ACTGAGAAGC GAAAAGAGAT TCATTGAGG AATGAAGCCT ATTATAAAAA ATTGTTAGAT	10320
AATTTTAAGG ACAAGGCCTA TATCACCTTG GCCACCTTGG ATGTTTCTAA ACCTTCGCAA	10380
GAGTTAGAAG AACAGTTAGC GAAAAATAGA GCCTTGGAAG AGACCTTTAC TGAGTCGACT	10440
CGAACTTCAA AAGTAGAAGC GCAGAAGAAG GAAAAAGAAC GTTTGTTAGA GGAATTGACC	10500
TTCTTGCAAG AATATATAGA TGTAGGTCAA GCGAGAGTTC CTTTAGCGGC TACTTTGAGT	10560
TTGGAATTTG GTACTACCTC TGTCATATA TATGCTGGTA TGGATGATGA TTTTAAACGT	10620
TACAATGCAC CAATTTTAAC TTGGTATGAA ACGGCTCGCT ATGCCTTTGA ACGAGGTATG	10680
ATCTGGCAAA ATTTAGGTGG TGTTGAAAC TCTCTCAATG GTGGACTTTA TCATTTTAAG	10740
GAAAAATTTA ATCCAACGAT TGAAGAATAC TTGGGTGAAT TTACAATGCC CACTCATCCT	10800
CTCTATCCTC TGTTAAGACT TGCTCTTGAT TTCCGTAAAA CATTAAGAAA AAAACATAGA	10860
AAGTAAGTAT ATGGCACTAA CAACACTCAC GAAAGAAGAG TTTCAGACTT ATTCTGATCA	10920
GGTTTCTTCT CGTTCCMTTA TGCAATCTGT CCAGATGGGG GATTTGCTAG AAAAAAGAGG	10980
GGCTCGAATT GTTTATCTTG CTTTGAAACA AGAAGGAGAA ATTCAAGTTG CAGCTCTGGT	11040
TTATAGCCTG CCCATGCTGG GTGGTCTGCA TATGGAACCT AATTCGGGGC CGATTTATAC	11100
CCAACAAGAT GCTCTTCCAG TTTTATATGC AGAGTTAAAA GAATATGCCA AGCAAAATGG	11160
TGTATTAGAG TTGCTTGTA AACCCTATGA AACTTATCAA ACTTTTGATA GCCAAGGTAA	11220
TCCAATAGAT GCTGAGAAAA AAAGTATTAT TCAAGATTTG ACTGATTTAG GTTATCAATT	11280
TGATGGCTTA ACAACAGGTT ACCCAGGTGG AGAACCAGAT TGGTTATACT ATAAAGATTT	11340
AACTGAATTA ACTGAAAAGA GTTTGCTTAA AAGTTTATAG AAAAAGGGTA AACCTTGGT	11400

473

GAAAAAGGCT GAAACCTTTG GCATTCGGTT GAAAAAGTTA AAACGTGAAG AACTATCGAT	11460
TTTAAAGAAT ATAACAAAAG AAACCTCTGA ACGTAGAGAA TATAGTGATA AAAGTTTAGA	11520
ATATTATGAG CATTTTATG ATACTTTTGG AGAACAAGCG GAGTTTCTCA TAGCAAGCTT	11580
AAATTTTTCG GACTATATGA GCAAATTGCA AGGTGAACAA AGTAACTAG AAGAAAACCT	11640
GGACAAGTTG CCACTTGATT TGAGTAAAA TCCTCATTCT GAGAAAAAAC AAAATCAACT	11700
GAGAGAATAT TCTAGTCAAT TTGAAACGTT TGAAGTTTGA AAAGCAGAAG CGCGAGACTT	11760
GATTGAAAA TATGGAGAAG AAGATATTGT TTTAGCTGGG AGTTTATTTG TTTATATGCC	11820
TCAGGAAACG ACTTATCTCT TTAGTGGTTC CTACACTGAG TTTAATAAGT TCTATGCCCC	11880
TGCACTGCTT CAAAAATATG TTATGTTGGA AAGCATAAAA CGTGAATAC CTAAATACAA	11940
CTTCCTAGGC ATTCAAGGGA TTTTGTATGG AAGTGATGGT GTTTTGCGTT TTAACAGAA	12000
TTTTAATGGC TATATTGTAC GCAAAGCAGG TACTTTCCGT TACCATCCAT CGCCTTTAAA	12060
ATACAAAGCT ATCCAGTTAC TCAAAAAAAT AGTAGGACGT TAAGATGAAA AAGTCAGTAT	12120
TTAGATTCTT TTTAGCTTCT TTTAGTAAAA TAATTCTTAT TTGCTAGAAA GGTGGAGAGA	12180
CATGCGCTGG CTTTTCGTT TGATAGGGGC TTTCTTTTCT TTTGTGTGGC GTTTGTTTTG	12240
GCGTCTGGTT TGGATAGTTG TGCTCTTATG TGTGCTTGCT TTCGGACTTC TCTGGTATCT	12300
GAACGGAGAT TTTCAAGGAG CGCTAAAGCA AGCAGAACGG TCAGTAAAAA TTGGTCAACA	12360
AAGTATTGAC CAATGGGAGA AAACAGGGCA ACTGCCTAAG TTAAGCCAGA CAGATAGTCA	12420
CCAGCATTCT GAAGGAAGGT GGGCAGAGC CTCTGCTCGT ATTTACCTGG ATCCGCAGAT	12480
GGATTCAAGC TTTCAAGAGG CTTATTTAGA AGCAATCCAG AACTGGAATC AAAGTGGTGC	12540
TTTTAACTTT GAACTCGTGA CTGAGTCTAG TAAGGCGGAT ATTACGGCTA CGGAGATGAA	12600
CGACGGAGGC ACTCCTGTGG CAGGAGAGGC GGAAAGTCAA ACTAATCTCT TAACAGGGCA	12660
ATTCCTGTCC GTAACGGTGC GGTGAATCA TTATTATTG TCCAATCCAT ACTATGGCTA	12720
CTCCTATGAA CGCCTTGTC ATACGGCAGA ACATGAGTTA GGTCATGCCA TTGGCTTGGA	12780
CCATACAGAT GAGAAGTCTG TCATGCAACC AGCAGGTTCC TTTTATGGTA TCCAGGAAGA	12840
GGATGTTGCA AACCTCCGAA AAATATATGA GACTAGTGAG TAGGGTACTA TCTTTCCCTA	12900
CTTTTTTTCG TATAATGGAA CTATGAACAA CTTGATTAAA TCAAACTAG AGCTCTTGCC	12960
GACCAGCCCT GGTGCTACA TTCATAAGGA TAAAAATGGC ACCATTATCT ATGTAGGAAA	13020
GGCTAAAAAT CTGCGTAATC GAGTACGGTC CTATTTTCGT GGAAGTCATG ATACCAAGAC	13080
AGAGGCTCTG GTGTCTGAAA TTGTGGATTT TGAATTTATT GTTACGGAGT CTAATATTGA	13140

474

GGCACTTCTC CTAGAAATCA ACCTGATCAA GGAAAACAAG CCCAAGTACA ATATCATGCT	13200
CAAGGATGAC AAGTCCTATC CTTTCATCAA AATCACCAAT GAGCGCTATC CACGCTTGAT	13260
TATCACTCGT CAGGTCAAAA AGGACGGAGG TCTTTATTTT GGACCCATATC CCGATGTGGG	13320
GGCAGCCAAT GAAATCAAGC GGTGCTGGA TCGGATATTC CCTTTTCGTA AGTGTACCAA	13380
CCCCCCCTCT AAGGTCTGTT TTTATTACCA TATCGGCCAG TGTATGGCCC ACACCATCTG	13440
TAAGAAGGAT GAGGCTTATT TCAAGTCTAT GGCCCAAGGAG GTGTCTGATT TTCTGAAAGG	13500
TCAGGATGAC AAAATCATCG ATGATCTCAA GAGTAAAATG GCAGTAGCAG CACAAAGTAT	13560
GGAGTTTGAA CGTGCGGCGG AATACCGTGA CCTGATTCAG GCTATTGGAA CGCTTCGAAC	13620
CAAGCAACGG GTCATGGCGA AAGATTGCA AAATCGCGAT GTCTTTGGCT ACTATGTGGA	13680
TAAGGGCTGG ATGTGTGTGC AGGTTTCTT TGTCCGTGAG GAAAGCTCAT CGAGCGCGAT	13740
GTCAATCTCT TCCCCACTT CAATGATCCA GATGAGGATT TTTTGACCTA TGTAGGACAA	13800
TTCTATCAAG AAAATCTCA TCTAGTCCC AATGAGGTAC TGATTCCGCA GATATTGACG	13860
AAGAAGCTGT CAAGGCTTTG GTGGATTCCA AGATTCTTAA GCCTCAACGT GGAGAGAAAA	13920
AACAACGGT CAATCTAGCC ATAAAAATG CTCGTGTTAG TCTAGAGCAG AAGTTCAATC	13980
TGCTAGAAAA ATCTGTCGAA AAGACTCAAG GAGCTATTGA AAATCTAGGG CGTTTGCTCC	14040
AAATCCCGAC CCCAGTACGT ATCGAGTCCT TCGATAACTC TAATATCATG GGAAC TAGCC	14100
CTGTTTCGGC TATGGTGGTC TTTGTCAACG GTAAACCGAG TAAGAAGGAT TACCGTAAGT	14160
ACAAGATAAA AACGGTTGTT GGACCAGACG ACTATGCCAG CATGAGAGAG GTCATTGCA	14220
GACGCTATGG TCGAGTACAG CGTGAGGCTT TGA CTCTCC AGATTGATT GTGATTGATG	14280
GGGGCAAGG TCAAGTCAAT ATCGCTAAGC AGGTTATCCA AGAGGAACTG GGCTTGATA	14340
TTCCAATGTC TGGGCTGCAA AAGAATGATA AGCACCAAAC CCATGAATTG CTCTTTGGAG	14400
ATCCGCTTGA GGTGGTGGAT TTGTCTCGCA ATTCTCAGGA ATTTTCTCTC CTCCAACGCA	14460
TCCAAGATGA GGTGCACCGC TTTGCTATCA CTTTCCACCG CCAACTGCGC TCCAAAATT	14520
CTTTCTCATC TCAATTGGAT GGGATTGACG GTCTGGGACC TAAACGCAAG CAGAACTTA	14580
TGAAGCATTT CAAGTCTTTG ACCAAAATCA AGGAAGCCAG TGTGGATGAG ATTGTGGAAG	14640
TTGGGTACC TAGAGTCGTT GCAGAGGCTG TGCAAAGAAA GTTGAACCCG CAGGGAGAAG	14700
CCTTGCCCTCA AGTAGCAGAA GAAAGAGTAG ATTACCAAAC GGAAGGAAAC CACAATGAAC	14760
CATAAAATCG CAATTTTATC AGATGTTTAT GGCAATGCGA CGGCGCTAGA AGCAGTGATT	14820
GCAGATGCTA AAAATCAAGG GGCCAGTGAA TATTGGCTTC TGGGAGATAT TTTTCTTCT	14880
GGTCCAGGCG CAAATGACTT AGTCGCCCTG CTAAGGACC TTCCTATCAC AGCAAGTGTT	14940

475

CGAGGCAATT GGGATGATCG TGTCCCTTGAG GCTTTAGATG GGCAATATGG CTTAGAAGAC	15000
CCACAGGAAG TTCAGCTCTT GCGTATGACA CAGTATTTGA TGGAGCGAAT GGATCCTGCA	15060
ACGATTGTCT GGCTACGAAG CTTGCCTTTG CTGGAAAAGA AAGAAATTGA CGGATTGCGC	15120
TTTTCTATCT CTCATAATTT ACCTGACAAA AACTATGGTG GTGACTTGCT AGTTGAGAAT	15180
GATACAGAGA AATTTGACCA ACTGCTAGAT GCGGAAACGG ACGTGGCAGT TTATGGTCAT	15240
GTTCACAAGC AGTTGCTTCG TTATGGAAGT CAAGGGCAAC AAATCATCAA TCCAGGGTCG	15300
ATTGGCATGC CCTATTTTAA TTGGGAGGCG TTAAAAATC ACCGTTCCCA GTATGCCGTG	15360
ATAGAAGTTG AAGATGGGGA ATTACTCAAT ATCCAATTC GTAAAGTTGC TTATGATTAC	15420
GAAGCTGAGT TAGAATTGGC CAAGTCCAAG GGGCTTCCCT TTATCGAAAT GTATGAAGAA	15480
CTGCGTCGTG ACGATAACTA TCAGGGGCAC AATCTGGAAT TATTAGCCAG CTTAATAGAA	15540
AAGCATGGGT ATGTAGAGGA TGTGAAGAAT TTTTTTGATT TTTTGTAAGA GTTTCCTAAA	15600
ATAGCCAATG CAAACTAAAA AAGCGATTTG CTGGTCCAAT CGCTTTTACT ATATCTTATA	15660
CTCAATGAAA ATCAAAGAGC AAAGTAGGAA GCTAGCCGTA GGTGCTCAA AGCACAGCTT	15720
TGAGGTTGCA GATAAAGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTGT TATTGTAACT	15780
GAGATTGATC TGGGAGCTAA GAACACCTA GATAGGTATT GCTGAGTTT TCAAGGGTTC	15840
CGTCTTGATA GAGTTCTTTG AGCGCTTTAT CAAATGCTC TTTAACTCT TTTTGGTCGC	15900
TTGAGAAAAT GATATAATTG CTGGGGCTAT CTGCAGAAGG TAAATCAACG ACTGAGAGGT	15960
CTAAACCACG GTCCTTGATA ATCTTTTGAA CGGATACCTT GTCAAAAAC AGGAATCAA	16020
ACTCTCCGTT AGCAAGGTCT AGGATTCGTT TACCAATATC CTCACCAGAA AAATTAATTG	16080
TAGCGGGATT ATCAGTGTGT TTCTGATTCC AGTTATTGAT GAATTGAGCG TTAGAAGTTC	16140
CGGTATCCTC TTGTGTTGTT TTACCAGCGA TCTGGTCAAG AGAAGTCAA GGATTTTCT	16200
TGTTGCTGAC AAGGACGAGG GGATTGTTGG AAATTGGAAG CGAGTAAAGG TATTTTTCAG	16260
CACGCTCTTT TGTGTAATC AAGTTATTGG CCGCAGCCTG ATAGTGACCA GAATCAAGTC	16320
CTGGGAAGAT GCTCTCCAG GCGGTTCTTT GGAATTGAAT CTCGTAGTCG CTGAGTTT	16380
CATCTACTGC CTTTAAAACT TCGATATCAA AGCCTGTCAG ATTGCCCTTG TCTTCGTAGT	16440
CAAATGGTGG CACGTCGCCA GCTGTAGCAA GGACGATTGT CTTTGTAGCG CTAGTCTCTT	16500
TGGGTGTAGC TTGATTCTCA CAGGCAACCA AAAATGGTAG GATAGCTAGT AATAGGCTAA	16560
ATTTTTTCAT ACTGTCTCCA TTCAAATGTA AAG	16593

(2) INFORMATION FOR SEQ ID NO: 53:

476

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3510 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

GGGATATCCT TATATCCTTG TTCCTGGAAC CATTTGTGGGA ATTGCTCAAC AGTTTTTTTCA	60
CCTTGAATTC CTGGTGCAAT GACAGTAAGA ATTTGCGAAAT CACGATCTGG TTTGCGCCGCT	120
AGTTCCATCA ACTCTGGCAT ACTTTTCTTG CATGGACCAC ACCATGAAGC CCAAACTTC	180
AAGTAAACCT TTTTACCCCT AAAATCAGAT AACTTAACTT CTTTGCCATC CATGGATTGC	240
AATGTGAAGT CTGGAGCATC TTTTCCAACA GCAATTTGTT GTACAGTCGT TTGTTGTTTT	300
GGCTGTTGTG CTGCTTGAGT CTTTCTAGTT TCTTCCTCAC CACAGGCCAT CAATACAAC	360
AATGACAAGA GACTTAAGCC AGCAAACATT ACTTTTTC TTTGTCCTCC TTTATTCAAA	420
AATTCAGCT AGAACATTTA CTTGTCCTAA TAGTAACAAA ATCCCATTA AAACAATGAG	480
GAAACCACCA ATTTTCTTTA GTAGCATCAT ATGACGCTTG ATTTTACTAA AATATGGCAT	540
GACTAGACCT GAAGCTAGTG CCAATACCAA GAAAGGAAG GCCATGCCaG AGTGTAATG	600
AGAGTATAAA TCGCTCCTTG CCAAGCGCCA TTGCCTCCAG AAGCCGCAAG TGCTAAAACA	660
GAACTTAAAA CTGGACCAAT ACAAGGTGTC CAACCAAAGC TAAAGGTAAT ACCAAGTAAA	720
AAAGCTGACC AATAACGATT AGAATCTGAT TTTTAAAGG TAAAACTTTT TTGAACCTCT	780
AATTTCTTCA AATGAAAAAT TTCCATCTGG TGAAGACCCA AAATGATAAT AATAGCTCCC	840
ATGCCATATC.GAAACCAATT TGCATAGAGA ATATGACCAA AGTAACCAGC ACCAAAGCCT	900
AGAATAAAGA AAATGAGAGA GATACCAGCG ATAAAGCAA GTGTTTGAAT CAAGCCTGAC	960
CAGAGAACCT TTCTCCCAA CAAAGAAAAG CTTTTTGCAC TTTCTTGATC ATCCAATAAA	1020
ATCCCAGCAT AGACTGGCAG AAGAGGAAAA ATACAAGGAG AAAAAAGGA TAAAACACCT	1080
GCTAGAAAAA CAGAGATTAA AAATACTATC GTTTCCAATA AAGAACCAAC TTTCTTAATA	1140
ATTCTAATCC TATTTTACTA TATTCATTT TATTTGTAAG CTTTCTGCTA CGCAAAATCG	1200
TATCGGGCAC TATTGGACCA ATCTTTTCTT TTGCTAGTCA AGGCGGATCT TATCCCCAA	1260
AATAGCCAAA AAGCAACGAC AAGGATTACT CATCGCTGCT TTTGTGAACG AAAATGCTTT	1320
TTAGGTCTGA CATTTCTATA ATCATGTTTT ACTTGAGTTT GTCAAGGATT GCTTTAAGCT	1380
CCTCTACTAG TTTAGTTTCT GTCTCTGCTG AGCCATTTTC TTCTTTCACG AAATCAAGGG	1440
TTTCTTGAG AAGGTTTTGG GCTTTGGCAA GGACTTTTTT ATCCGCTTTT TCTGCATCTA	1500

477

GCTGTCCTAG AACCTTGATC AATTCCTGTC TTAATTGCTG GATTTCTGAC TCTTCTTAC	1560
GGCGAATCAG CCAGAAGGCA ATCACGCCTA GGAGGGCAAG TAGACTGACC ACAATCACTC	1620
CTGCCGGAAC TGAGTTTGT TCAGTCATCT TATCTGAATC CTTACTATCT TCCGTTCCCTT	1680
GTTTTGCATC CTCTTGTCC TGTGCAGGCT TGCTGTCGCT AGCATTTGCT TFCACATCTT	1740
TGAGAGAGTC CAAGGCAGCC CAGCCTTCAC AGACTCTACT GCAGTATGCA GACCTTACTC	1800
TGTCAAGGCA CTATCTCCG GAGCTTTTG AGCATCTAGG AGGACAGCCT TGTTTGCATC	1860
GATTTTCGGA TCAGATACTG TTGCCAAAGC TTTCAAGCGT TGGTCTAACT CTTGACTCAA	1920
GGCAGGAAGT TCAGACTTGT CAACTTGCTC TTGAGCTTGT GTGCTCGTTG AGCTAGCCGA	1980
AGCGCTTGCT ACCACTCTAG GATCTTGAGT CGGAGCTGAG CTTGGAGCTG GGACAGGGCT	2040
TGCAGGTGA CTAGGAACAG TTATGGTATA TTGAACTAG AATAGTACAT ATGGACTTCT	2100
AAAACATTGT TAGAATTCGA TTTTACTGTC CTGATCGATT TGTCTATTTC TTATTTTATT	2160
TTACTATAAT AACCGATGGT GTGGTTAATG TTGGTAAGAG AAACCTCTGA AACCAAGCTT	2220
CAAAAAAGTC GCTCGTCATC GTCTCTCGT AAGTCATTGG AGCGATTAAAT TCACCATTTG	2280
TTAGACCTGC AACCAAAGAA ATCCTCTGAT ATCTTCTTCC AGATACTTTG CCTCTTATTA	2340
ACTGACCTTT TAATGAGCGA CCATATTCTC GATAAAAAATA AGTATCGAAT CCTGTTTCGT	2400
CAATCTAAAC AGGTGCTAGG TGCTTTAAAC TATTAAATTT CTTAAGAAAT AAGGCTACTT	2460
TTTCTGGGTC TTGTTTATAG TAGGTGTGGT TCTTTTTC GAGTGTAGCC CATAGCTTTG	2520
AGCGCATAGT GGATGGTAGT TGGATGACAG CCAAATCAG AAGCTATTTC AGTCAAATAA	2580
GCCTCTGGAT TGTCAGTAAG ATAGTTTTTA AGTCTATCTC TATCAACTTT TCTTGGTTTT	2640
GTTCCTTTTA CTGGTGGTT TAGCTCTCCT GTTTTCTCTT TTAGCTTTAA CCAGCCATAA	2700
ATGGTATTAC GTGAGATTG GAAAACGTGT GATGCTTCTG TTATACTACC TATTCGCTCA	2760
CAATAAGAGA GAACTTTTTT ACGAAAATCT ATTGAATATG CCATAAGAAG ATTATACCAC	2820
ATTGTGTAAT ATTTTGGTT CATTTCACTA TAACACAAAA TAGATTATTA TTACATAACA	2880
AAAAAGAGGT CTAAACCTCT TAACTCAATT ACTCCGCCAG TAGGACTCGA ACCTACGACA	2940
TCATGATTAA CAGTCATGCG CTACTACCAA CTGAGCTATG GCGGATTAAA GCTAAGCGAC	3000
TTCCCTATCT CACAGGGGGC AACCCCAAC TACTTCCGGC GTTCTAGGGC TTAACCTCTG	3060
TGTTCCGCAT GGGTACAGGT GTATCTCCTA GGCTATCGTC ACTTAACTCT GAGTAATACC	3120
TACTCAAAAT TGAATATCTA TTCAATTTAA GAAAACCGTT CGCTTTCATA TTCTCAGTTA	3180
CTTTGGATAA GTCCTCGAGC TATTAGTATT AGTCCGCTAC ATGTGTCGCC AACTTCCAC	3240

478

TTCTAACCTA TCTACCTGAT CATCTCTCAG GGCTCTTACT GATATATAAT CATGGGAAAT	3300
CTCATCTTGA GGTGGkTtCA CACTTAGATG CTTTCAGCGT TTATCCCTTC CCTACATAGC	3360
TACCCAGCGA TGCCTTTGGC AAGACAACTG GTACACCAGC GGTAAGTCCA CTCTGGTCCT	3420
CTCGTACTAG GAGCAGATCC TCTCAAATTT CCTACGCCCG CGACGGATAG GGACCGAACT	3480
GTCTCACGAC GTTCTGAACC CAGCTCGCGT	3510

(2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20986 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CGGAGAAAA CATGGCTAAG TCAAACTTTG AAAAAGTAGA ATCAGTTGTT GGCTGGGTTC	60
GTGATAAGAA AATCACAGGC TACCGTATCT CTAAGAAAC CAATGCGCGT GAAATGTCTA	120
TCATTGCTCT GCGCAGGGT CGTGCAAAG TAAAAATAT TTCATTGAA ACAGCCCTAG	180
GCCTAATTGA TTTCTATGAA AAAAATTATG AAAAATTGA AGATTAATCT TTGATAACG	240
GCGGATTCTT GACCTTCAAG TAGTAGAGAT AGAGAATCTG CCTTTTCATT TTGAGGACAG	300
CAAAAAGACT GCACGGTTGA TGCAGCCTTT TCTTTTATT TGAGATAGCG TTGAAGGAAC	360
TCTTTTGTTT GGTCTTCTT AGGATTGGTG AAGAGTCTT CTGGTTTACC TTCTTCAGCG	420
ATCACGCCCT TATCCATAAA GATAACACGG TGAGAGACAT CACGGGCAA TTCCATTTC	480
TGGGTTACGA CAATCATGGT CAAGCCTTCC TGAGCCAGGT CCTGCATGAT TTTGAGGACT	540
TCTCCAACCA TTTCTGGATC GAGAGCTGAT GTTGGTTCAT CAAAGAGAAT AGCGTCCGGA	600
TTCATGGAGA GGGCACGAGC GATGGCCACA CGTTGTTTTT GACCACCTGA GAGTTGTTTT	660
GGTTTGGCTT GCCAGTAGCG TTCTCCCATG CCGACCTTTT CCAGGTTTTT TTTGGCAATC	720
TTTTTCAGCTT CTGTGCGTTC GCGTTTATAG ACAGTTGTCT GAGCGACGAT TGTGTTTTCA	780
AGAACATTGA GATTTTCAA GAGGTTAAAG GATTGGAAAA CCATCCCCAA CTTTTCACGG	840
TATTGCGTGA GGTTCATAGC TTTTTCGAGG ACGTTTTGTC CATGATAAAG GATTTGTCCA	900
TCAGTTGGTG TTCAAGTAG GTTAATGGAG CGTAGGAAGG TCGATTTTCC GCTTCCAGAG	960
CTTCCGATGA TAGAGATGAC CTCTCCCTTG TGGACAGTGA GTGAAATGTC TTTTAGCACT	1020
TCGTTTGTG CATAGGATTT TTTGAGGTGT TTAATTTCAA GGATTGCTTG TGTCATTATT	1080
TCAAATCCTC CGTTTGCAAT TGGTTAGCAC CTGTAGTGA GGTATCCATG TCCATTCTGC	1140

479

GCTCGATAAA GCGTAGGATA CGTGTTACGG TGAAGGTGAG GACAAAGTAA ATCACGGCGA	1200
TGATTGTAAA TGTCTGGAAG TATTGATAGG TTTGTGTTGC CACGGTATTT CCTGAGAAAT	1260
AAAGTTCGAC AACAGAGATA ACGTTCAATA CAGATGTATC TTTGATATTG ATGACAAATT	1320
CATTACCAGT TGCAGSTAGG ATGTTACGGA CTACCTGAGG TAGGACAATC TTACGCATGG	1380
TCTGGTTATG GGTACATACCA AGAGCAGTCG CAGCTTCAAA TTGTCCCTTG TCAACTGCTA	1440
GGATACCACC ACGGACGATT TCAGTCATGT AGGCACCGGT ATTGATTGAA ACGATGAAGA	1500
TAGCAGCCAG TGTACGGTCA AGGTGATCC CGAAAGCTTG GGCAGTTCCA TAGTAGATAA	1560
CCATCGATTG AACAAATCATT GGCCTACCAC GGAAAATTTC AATGTAGACA TTGAGAACCC	1620
AGCCGACTAG TTTTGTAGG CCGTAAATGA CTTTGTTC AGAGAGAGGA GCAGTACGGA	1680
AGACACCAAT GGCAAGTCCA ATAATGAGAC CTATGATGGT TCCGACGATA GAGATTAAAA	1740
GAGTGATACC AGCACCACGC AAGAGTTGTT GCCAGTTTC AGAAAGAATT TTAGCAACTT	1800
GGCTAAAGAA ACTACTGCTA GTCTCTCAG TTGTTGTAGC TTCGGCAGGT TGTTCCTGA	1860
TCATACGATC CATCAAGGCA ACTTGGTCAT CTTTGTAAAT GTTTCAATG CTGGCATTGA	1920
TTTGGCTAAT ACGATTGTCA TTTTACGAA GCCCAGTAGC GATAGCTGTA TCTTCTCCC	1980
CAGTTTGTAA ACCAGGTTCT ACTTGAATCA TCTTGAAGTT AGAGTTCGCA GCTTCAGCAG	2040
TCAGTGCTTC TGGACGTTCA GAAACATAAG CATCAATGAC ACCAGCCTCA AGAGCTGTG	2100
GCATTTGAGC GAAGTCTCCC ATGGCTGTTT CTTTTTAGC ACCTGGGATT TGTGCAATCA	2160
AGTTATAAAG GTAGACCCCT TGTTGAGAAG TGATTTTTC ACCGTAAAG TCATCCAAAG	2220
ATTTAGCACT TGCCTAGGCA GAATCTTTT TGACAAGCAA AACTGGTTCG CTAGTATAGT	2280
AACTGCTCGA AAAGGCAATT TCTTGTTC GTTCTGCAGT TGGACTCATA CCTGCGATAA	2340
TCATGTCAAT CTTACCAGAA GTAAGGGCAG GGAAGTAGC TTCCCACTTG GTTTTAACAA	2400
CCAAAGGTTT TTTACCTAAG TCCTTAGCGA TTTCTTGGC GATTTGAACA TCGTATCCGT	2460
TGGCATACTG ATTGGTCCCA TCGATTTTGA CAGCTCCGTT GCTATCATCA TCCTGGGTCC	2520
AGTTAAAGGG AGCATATGCT GCTCCATAC CGATGCGTAA ATATTTCATCG GCTTGAGCAA	2580
CATTGACAAG TCCTAGCATC AGCAAGAGAC TTGTGAAAAT AGATAAGTAY ATGTGGCTCA	2640
TGATTTCTCC TATTCTGATC TATTAAAAA TAACTGTCTC CTATTTTATC GAAAAATGCC	2700
TAATTTTCA ACATAAGTAA GTCTTTACTT ACGAAAAAT GCTATAATGA TAAGAAAGAT	2760
AAAAAGGGG CTTAGTTGAT GAAAAAACT TTTTCTTAC TGGTGTTAGG CTTGTTTGC	2820
CTTCTCCAC TCTCTGTTT TGCCATTGAT TTCAAGATAA ACTCTTATCA AGGGGATTTG	2880

480

TATATTCATG CAGACAATAC GGCAGAGTTT AGACAGAAGA TAGTTTACCA GTTTGAGGAG	2940
GACTTTAAGG GCCAAATCGT GGGACTTGGA CGTGCTGGTA AGATGCCTAG CGGGTTTGAC	3000
ATTGACCCTC ATCCAAAGAT TCAGGCCGCG AAAAACGGTG CAGAACTAGC AGATGTGACT	3060
AGCGAAGTAA CAGAAGAAGC GGATGGTTAT ACTGTGAGAG TCTATAATCC AGGTCAGGAG	3120
GGCGACATAG TTGAAGTTGA CCTCGTCTGG AACTTAAAAA ATTTACTTTT CCTTTATGAT	3180
GATATCGCTG AATTAAATTG GCAACCTCTG ACAGATAGTT CAGAGTCTAT TGAAAAGTTT	3240
GAATTTTCATG TAAGGGGAGA CAAGGGGGCT GAAAACTCT TTTTCCATAC AGGGAACTT	3300
TTTAGAGAGG GAACGATTGA AAAGAGTAAC CTTGATTATA CTATCCGTTT AGACAATCTT	3360
CCGGCTAAGC GTGGAGTTGA GTTGCATGCC TATTGGCCTC GGACCGATTT TGCTAGCGCT	3420
AGGGATCAGG GATTGAAAGG GAATCGTTTA GAAGAGTTTA ATAAGATAGA AGACTCGATT	3480
GTTAGAGAAA AAGATCAGAG TAAACAACCTC GTTACTTGGG TCCTCCCTTC GATCCTTTCC	3540
ATCTCCTTGT TATTGAGTGT CTGCTTCTAT TTTATTTATA GAAGAAAGAC CACTCCTTCA	3600
GTCAAATATG CCAAAAATCA TCGTCTCTAT GAACCACCAA TGGAATTAGA GCCTATGGTT	3660
TTATCAGAAG CAGTCTACTC GACCTCCTTG GAGGAAGTGA GTCCCTTGGT CAAGGGAGCT	3720
GGAAAATTCA CCTTTGATCA ACTTATTCAA GCTACCTTGC TAGATGTGAT AGACCGTGGG	3780
AATGTCTCTA TCATTTTCAGA AGGAGATGCA GTTGGTTTGA GGCTAGTAAA AGAAGATGGT	3840
TTGTCAAGCT TTGAGAAAGA CTGCCTAAAT CTAGCTTTTT CAGGTAAAAA AGAAGAACT	3900
CTTTCCAAIT TGTTTGCGGA TTACAAGGTA TCTGATAGTC TTTATCGTAG AGCCAAAGTT	3960
TCTGATGAAA AACGGATTCA AGCAAGAGGG CTTCAACTCA AATCTTCTTT TGAAGAGGTA	4020
TTGAACCAGA TGCAAGAAGG AGTGAGAAAA CGAGTTTCCT TCTGGGGGCT CCCAGATTAT	4080
TATCGTCCCTT TAACTGGTGG GGAAAAGGCC TTGCAAGTGG GTATGGGTGC CTTGACTATC	4140
CTGCCCCCTAT TTATCGGATT TGGTTTGTTT TTGTACAGTT TAGACGTTCA TGGCTATCTT	4200
TACCTCCCTT TGCCAATACT TGGTTTCTA GGGTTAGTTT TGTCTGTTTT CTATTATTGG	4260
AAGCTTCGAC TAGATAATCG TGATGGTGTT CTAATGAAG CGGGAGCTGA GGTCTACTAT	4320
CTCTGGACCA GTTTTGAAAA TATGTTGCGT GAGATTGCAC GATTGGATCA GGCTGAACTG	4380
GAAAGTATTG TGGTCTGGAA TCGCCTCTTG GTCTATGCGA CCTTATTGG CTATGCGGAC	4440
AAGGTTAGTC ATTTGATGAA GGTTCATCAG ATTCAAGTGG AAAATCCAGA TATCAATCTC	4500
TATGTAGCTT ATGGCTGGCA CAGTACGTTT TATCATTCOA CAGCACAAAT GAGCCATTAT	4560
GCTAGTGTCTG CAAATACAGC AAGCACCTAC TCTGTATCTT CTGGAAGTGG AAGTTCTGGT	4620
GGTGGCTTCT CTGGAGGCGG AGGTGGCGGC AGTATCGGTG CCTTTTAAAG AGAGCTACCA	4680

481

TAGACTGAAA AAGTATGATA TAATGGAAGA TAGAAAAAG ACAAACTATA AGAAAAGTCA	4740
ATAGTTTTAT CTAAACTATT TCTTATTTC AATTGATGAT TTGGCGATGA TTTTAGAGCA	4800
CGGCAAAAAG CCCTTGAAAA AGTCCATTTT TTCAAAGGTA ATCCTGTGTT AATTTAGAA	4860
ATTACATCAC TTTTGTTCG TCAAAATGGCA GCTCTTTTTT AGGATATAAA ACAGGGTTCG	4920
GATAAGTTTT TTTGCAAGGT GGATGATGGC TACATTGTAA TGTTTTCTT ATTCTAACTT	4980
AGTCTTAAGA TAGGCCTTAG AAGCAGGTGA AAAGCGAGGG CATGCTTTGG CAGCTTGAT	5040
GAGTGCCAC CGCAGATGAG GGAACCCCG TTGACCATT CTCCAGCTA AATCAATCTG	5100
ACCTGACTGA TAAATAGAAG AATCCAGTCC AGCGAAAGCT TGTAATTGAG CAGGATTATC	5160
AAAGGCATGA ATATTTGAA TCTCGGCTAA AATGACCGCC CTAAACGATC CCCAATCCCA	5220
GTAACCGTCG TGATACCGA GTTGAATCA GCCATCGAGT CATGATACA TGTTCCGCC	5280
TTGTCAATGA GCCTCTGTG ATGCTTGATG ATTTGCAATT CACGAGCAGG AGATGTTGTT	5340
CCGATAGAAC GAGGTGCGAC TGAGAGGATA TCCTGAATTT TAGAAGCGGT CAATCGCTTA	5400
ATTTCTATCA GCTTATCAA TCCTGCCTCA ATCCTTTTCT GAGGATTAGG GTAGCGTGC	5460
AAGAGTTGGT AGGTATATC TGAATGCTTT CCAACGATTT TATCCAATC AGGAAAGATG	5520
ATATCAAGAC AACGAGTGTA TTGTACTTTC CAATCAGACT GTTTTCTTG AGACGATGAA	5580
TATGCTAGC CAGTATTTT AGGTCTACTT GCCGATTATC GTGTGAAAT TGTTACGAT	5640
TGGGTCAGA AAGAAGTTA AGAGCGATGC CATGAGCGTC TTTCTTATCC GTTTTAGTCT	5700
TGCGAAGTGA TAATGATTTG GCAAATCCT TGATGAGCA AGGATTGTAG GTGTAACTT	5760
TATATCCTTG TTCATGCAGG AAGTTCAGTA GATTAAAGGC ATAATGTCCA GTATCTTCAA	5820
GAGCGATGAG ACAGTCTTG TTGATCTGTC GAATAGACAG ATCTAAGAGT TCAAAACCAG	5880
CTTATTATT TGAAAAAGTG AGTGGTTTAA GAACAGTTT TCCTGGAACA TTCAAGGCTG	5940
TAACATCGTG TTTATTTT GCGATATCAA TGCCTACATA AAGCATGGGA GTACCTCCAG	6000
ATATAGTATT TCAAGTCTAC TTGGTTATCC ACGAATTTT TGCCTTGTTA CCTTAGACGA	6060
GATCAAAAGT CTATGCGTTA TCAAACTCAT TACCAATTGA AACAAAAGCT GTGGTTAGAG	6120
CCTTTCGGAA ATCGTCAAGC GATTGGAGGA AATGAATAA TCCATAGTGG CTTATTCCAA	6180
GTATACCACT TGGGCTTTG CAGTAGCTAA CTGCGCTAAA TATAATATAG GGAGTAATCT	6240
ATGTATCTTA TTGAAATTTT AAAATCTATC TTCTTCGGAA TTGTTGAAGG AATTACGGAA	6300
TGGTTGCCGA TTCCAGTAC AGGTCACTTG ATTTTAGCAG AGGAATTCAT CCAATACCAA	6360
AATCAAAATG AAGCCTTTAT GTCCATGTTT AATGTCGTGA TTCAGCTTGG TGCTATTTA	6420

482

GCAGTTATGG TGATTTATTT TAACAAGCTC AATCCTTTTA AACCGACCAA GGACAAACAG	6480
GAAGTTCGTA AGACTTGGAG ACTATGGTTG AAGGTCTTGA TTGCTACTTT ACCTTTACTT	6540
GGTGTCTTTA AATTTGATGA TTGGTTTGAT ACCCACTTCC ATAACATGGT TTCAGTTGCT	6600
CTCATGTGTA TTATCTACGG GGTTCCTTC ATCTATTTGG AAAAGCGCAA TAAAGCGCGT	6660
GCTATCGAGC CAAGTGTAA AGAGTTGGAC AAGCTTCCTT ATACGACCGC TTTCTATATC	6720
GGACTCTTCC AAGTCTTGC TCTTTTACCA GGGACTAGCC GTTCAGGTGC AACGATTGTC	6780
GGTGGTTTGT TAAATGGAAC CAGTCGTCA GTTGTGACAG AATTTACCTT CTATCTTGGG	6840
ATTCTGTGTA TGTTTGGAGC TAGTGCCTTA AAGATTTTCA AATTTGTGAA AGCCGGAGAA	6900
CTCTTGAGCT TTGGGCAATT GTTTTGTCTC TTGGTCGCGA TGGGAGTAGC TTTTGCAGTC	6960
AGCATGGTGG CTATTCGCTT CTTGACCAGC TATGTGAAA AACACGACTT CACCCTTTTT	7020
GGTAAATACC GTATCCTGCT TGGTAGTGT TTGCTACTTT ACAGTTTGT CCGTTTATTT	7080
GTATAAGAAA AACCTGAAG GGGCAACTCT TCAAGGTTTT ATACTCTTCG AAAATCTCTT	7140
CAAACCGCGT CAGCTTTATC TGCAACCTCA AACAGTGTT TTGAGCAGCn CTGCGGCTAG	7200
CCTCCTAGTT TGCTCTTGA TTTTCATTGA GCTTTAAAT CCAGTCATGG TAATCCCCAA	7260
TAGGCGGACA CCTCTTCTT TCTTGCTTAA TTCTTCATAG AGTTGCAGGG CTATTTGGCT	7320
TATCTGACTA GCATCTTGTG TTTTGTGAGC AAGACTTTTT CGTTTGGTAA GAGTTGAAAA	7380
GTCTCTGTAG CGGATTTTCA AAATGACAAT TTTTCCAGCT TTTTCTTGT GATGTAGATT	7440
GAGAGCGACT TTTTCTGATA GAAGAGTCAG CTCTTTTTTG ATATCTTCCT CAGCAAGGAG	7500
AATCTTCCCG TAGGTTTCTT CCTTGCCGAT TGATTTACGG ATGCGATTGG ATTTGACTGG	7560
AGAGTTGTGA ATGCCACGAG CCTTTCGATA CAGATCATAG CCTAGTCTAC CAAAACGGTC	7620
TATTAGGGTT ACCTCAGGAA CTTCAAGTAA ATCAGCACCA GTAAAAACGC CCATTTGATG	7680
AAGACGTTCT ACTGTCTTTT TTCCTACTCC ATGAAATTTG GAAATATCCA TTTGTTTGAG	7740
AAAAFCCTCA GCCTGTTTCA GTAGAATCAC TGTCAAACCA TGTGGTTTTT GATAATCACT	7800
CGCCATTTTA GCTAAGAATT TGTGTGAAGA AACGCCTGCG GAAGCAGTTA GATGGAGTTC	7860
TTGCCAGATA TCTTTTGA TAAGGCGAGC AATTTTGACC GCTGACTTGA TACCGAGTTT	7920
ATTTTCTGTC ACATCCAAAT AGGCTTCGTC AATGCTCATG GGTTCAATCA AATCTGTATA	7980
GCGCTTAAAA ATAGCTCGAA TCTGGAGTCC CACAGACTTG TATTTCTCAT AATTCCTGA	8040
GATAAAGACA GCCTGGGGAC AACGTTTATA AGCTTCCTTG GAACTCATGG CAGAATGGAC	8100
ACCAAAAGCT CTGCTTCAT AACTACAGGT AGAAACGACT CCCCCTCCAC CTGTTTGCCG	8160
AGGGTCGCTT CCAATAATGA CAGGTTTTCC TCTGAGTTTA GGATTATCCC TGATTTCCAC	8220

TGCAGCAAAA AAGGCATCCA TGTC AATATG GATGATTTTT CTGACAAAT CATTTAACAA	8280
AGGAAAAATC AACATGCCTA GCACCTTTTT ATACTCTTCG AAAATCTCTT CAAACCACGT	8340
CAGCTCTATm TGCAACCTCA AAACAGTGTT TTGAGCAATC TCGCGCTAGC TTCCTAGTMT	8400
GCTTTTCGAT TTCCATTGAG TGTTACTGCT TATTyTCTTT TATTATACCC TTTTTCCTGA	8460
AAAAAAGAAA AAAGGACTTT ATTTTTTCAA AAATATAATA CAGTTTGAAA TAAATATAG	8520
ACTGTTTTAG AAAAGAAAGT GTAAAAATAG GGAATTTTCA CTGTTGAAA TCGGTTACTA	8580
TATGGTATAC TTGTCTTATG AATGTAACAG ATGACTGTTA CTAGAAAAAA GAGGACATTA	8640
ATATGGTTGT TAAGACAGTT GTTGAAGCAC AAGATATTTT TGACAAAGCT TGGGAAGGCT	8700
TCAAAGGCGT AGATTGAAA GAAAAAGCAA GTGTATCAG ATTTGTACAA GCTAACTACA	8760
CACCTTATGA TGGAGACGAA AGCTTCCTTG CAGGACCAAC AGAGCGTTCA CTTACATCA	8820
AGAAAATTGT AGAAGAACT AAAGCACACT ACGAAGAAAC TCGTTTCCCA ATGGACACTC	8880
GTCCAACATC TATCGCTGAT ATCCCTGCTG GATTATCGA CAAAGAAAAT GAAGTTATCT	8940
TCGGTATCCA AAACGATGAA CTCTTCAAAT TGAACCTCAT GCCAAAAGGT GGTATCCGTA	9000
TGGCTGAAAC TACTTTGAAA GAAAATGGAT ACGAACCAGA CCCAGCTGTT CACGAAATCT	9060
TCACTAAATA TGTAACAACA GTTAACGACG GTATTTTCCG TGCCTACACT TCAAATATTC	9120
GTCGCGCTCG TCACGCACAC ACTGTAAGTG GTCTTCCAGA TGCATACTCA CGCGGACGTA	9180
TCATCGGTGT TTACGCACGT CTGCTCTTT ACGGTGCAGA CTACTTGATG CAAGAAAAAG	9240
TAAATGACTG GAATGCAATC AAAGAAATCG ATGAAGAAAC AATCCGTCTT CGTGAAGAAG	9300
TAAACCTTCA ATACCAAGCA TTGCAACAAG TTGTTCCGCT GGGTGACCTT TACGGGGTTG	9360
ATGTTTCGAA ACCAGCGATG AACGTGAAAG AAGCAATCCA ATGGGTTAAC ATTGCTTTCA	9420
TGGCTGTCTG CCGTGTGATT AACGGTGCTG CTACATCTCT AGGTCTGTGA CCAATCGTAT	9480
TGGACATCTT TGCAGAACGT GACCTTGCTC GTGGTACATT TACTGAATCA GAAATCCAAG	9540
AATTCGTTGA TGATTTTCGTT ATGAACTTC GTACAGTTAA ATTTGCTCGT ACAAAGCTT	9600
ATGACCAATT GTACTCAGGT GACCCAACCT TTATCACAAC TTCTATGGCT GGTATGGGTA	9660
ACGACGGTCG TCACCGTGTT ACTAAGATGG ACTACCGTTT CTTGAACACT CTTGACAACA	9720
TCGGTAACTC ACCAGAACCA AACTTGACAG TTCTTTGGAC TGACAAATTG CCATACAAC	9780
TCCGTCGCTA CTGTATGCAC ATGAGCCACA AACACTCTTC TATCCAATAC GAAGGTGTAA	9840
CAACAATGGC TAAAGACGGA TATGGTGAAA TGAGCTGTAT CTCATGCTGT GTGTCTCCAC	9900
TTGATCCAGA AAATGAAGAA CAACGCCACA ACATCCAGTA CTTTCGGTGCT CGTGTAACG	9960

484

TTCTTAAAGC CCTTCTTACT GGTTTGAATG GTGGTTACGA CGATGTTTAC AAAGACTACA	10020
AAGTATTTGA TATCGAACCA ATCCGTGACG AAGTTCTTGA ATTTGAATCA GTTAAAGCGA	10080
ACTTTGAAAA ATCTCTTGAC TGGTTGACTG ACACCTTACGT AGATGCCTTG AACATCATCC	10140
ACTACATGAC TGATAGGTAC AACTACGAAG CTGTTCAAAT GGCCTTCTTG CCAACTAAAC	10200
AACGTGCCAA CATGGGATTC GGTATCTGTG GATTTGCTAA CACTGTTGAT ACATTGTCAG	10260
CTATCAAATA CGCTACAGTT AAACCAATCC GTGACGAAGA TGGCTACATC TACGATTACG	10320
AAACAATCGG TGACTIONCA CGCTGGGGTG AAGATGACCC ACGTTCAAAC GAATTGGCAG	10380
AATGGTTGAT CGAAGCTTAC ACAACTCGTC TACGTAGCCA CAACTATAC AAAGACGCAG	10440
AAGCTACAGT ATCACTTTTG ACAATCACAT CTAACGTTGC TTACTIONAAA CAACTGGTA	10500
ACTCACCAGT TCACAAAGGT GTATACCTCA ACGAAGATGG TTCTGTGAAC TTGTCTAAAC	10560
TTGAATTTCT CTCACCAGGT GCTAACCCAT CTAACAAAGC TAAAGGTGGT TGGTTGCAAA	10620
ACTTGAATC ACTTTCTAGC CTTGACTTTA GTTATGCAGC TGACGGTATC TCATTGACTA	10680
CACAAGTATC ACCTCGCGCT CTTGGTAAGA CTCGTGATGA ACAAGTTGAT AACTTGGTAA	10740
CAATTCTTGA TGGTTACTTC GAAAACGGTG GACAACACGT TAACTTGAAC GTTATGGACT	10800
TGAACGATGT TTACGAAAAA ATCATGTGAG GCGAAGACGT TATCGTACGT ATCTCTGGAT	10860
ACTGTGTAAA CACTAAATAC CTCACTCCAG AAAAAAAC TGAATTGACA CAACGTGTCT	10920
TCCACGAAGT TCTTTCAATG GATGACGCCT TGGATGCATT GAGCTAATCA AGTTCTTGAA	10980
TAATAAAAAG GAACCTCGG TCAAACGACT GAGGGTTTTG TGCTTGGGAT AGTATGAGCA	11040
ATTCTTTCGG CGCAATATGC AATGTTTTTG GGCTCTTTGT CAACTGTAGT GGGTTGAAAA	11100
AAAGCTAAGC TTGAGAAAGG ACAAATTTG TCCTTTCTTT TTTGATGTTT AGGGCGATAA	11160
AAATCCGTTT TTTGAAGTTT TCAAAGTTCC GAAAACCAA GGCATTGCCG TTGATGTCTT	11220
TGATGAGTTT GTTAGTGGCC TCAAGTTTAG CGTTAGAATA AGGCAATTCA ATGGCGTTAG	11280
TGATGTAGTT TTTATAGCAA ATAAATGTGC TCAAAGTGGT TTAAAGGTG CGGTTGAGAT	11340
GAGGTAACGT GTCTTGAATT AAGCCCCAAA ACTGGTCAGT ATTCTTCTCT TGTAGATGAA	11400
ATAGGAGTAG TTGATACAGG TCATAGTAAT CTTAAGTTT AGGTACTAGA GTAAAGATT	11460
TCTTCAGACA CTCCTAGGA GTTAAGGTCT CTCTGAAAGT TCTAGCATAG AAAGGCTTAA	11520
GAGAGAGTTT CCGACTATCT TTTAGGATAA ATTTCCAGTA ATATTTAAGA GCTCTGTATT	11580
CCAGAGATT ATCATCAAAT TGCTTCATGA TGTGATTCT AGTCTGATTA AGAGCCCTGC	11640
TCATGTGTTG GACAATGTGG AAACGATCGA GAACAATTTT AGCATTGGGA AATAATTTCT	11700
TAATGAGAGG GATATAACTT CCAGACATAT CAACAGTGAC GACTTTAACT TTTTCTTAG	11760

CTTCTTTTCA GTACTTGAAG AAATGATTTT GGATGGTTGT TTGACGCTCTG TTATCAAGAA	11820
TGGTCATGAT TTTCTTAGTG TTGAAATCCT GAGCAATGAA AGCCAATTTT CCCTTCTGGT	11880
AGGAGAATTC ATCCCAGGAG AGGATTTTCAG GCAAAGTGGT GTAATCCTCT TGGAAATGAA	11940
ATTGCTTGAG CTTACGATAG ACGGTAGAGG TAGAGGTAGA GGTAGAGATG GCTAATTTAG	12000
AAGCGATATG TGTAAGAGCC TCTCTGTTGA GTAGGAGTTG GGCAATTTTC TGTCTCACCA	12060
TTTCCGAGAT TTGGCAATTT TTCTGAACGA GAGTTGTTTC AGCTACAGTG ACTTTCCGAC	12120
AGGACTTGCA TTGAAATCGT CTCTTTTTC AATGAATGAG GCTAGGGAAA CCACCAATCT	12180
CGATAAAAGG GATTTTAGAA GGCTTTTGGG AGTCGTATTT GATTGTGTTT CCTTTACAGT	12240
GTTTACATTT AGGTGGGTGA TAATCAAGTG TAGCGAAGAC TTCGATATGG GTATCGTGCT	12300
GAATGGCTTT ATTTAAGGTG ATGTTTTTGT CTTTTATTCC GATGAGTAAT GTGGTATGAT	12360
TGATGTGTTT CATAAGATAC TTTCTAATGA GTTGTTTAGG CGCTTTTCAT TATAAGTCTT	12420
ATGGGACTTT TTTGATACCT AAAAAGCCCT ATAATCTCCA CAGTGGGATT TACCCACTAC	12480
AGAAATTATA GAGCCAGAAA AAACACTTTT GTTCACTAGC AGAAACTAGA GAGCAGAAGT	12540
GTTTTTCTGT TCAGATTTAC CCAAACCTGG GAAATATGGG GATAAGAATA GAGATGGCTT	12600
AGGAAGCCCC TTTTTGTGTG TAGACAGTAC GATGAACTTA TAACAAATAG TGAGCCTTTT	12660
TAGCAATCAT TGCGACCCGT TTGTCAAAAG CCTCTMTTCG GATATCTACA ATTGTCTGAT	12720
AGATGAGACG CTGTTGGCTA ACATGCAAAAT CTAAGGCAAT CGTCAAAAAG TGATGTTTCC	12780
CTTTGGGATA CTGCTTTTTA ACGTAAGGCA GGTATTCTTT CGTTGTAATA ATAATCAATG	12840
GCTCTGTCAA ATGCTCCTCT GAAGGAGGAG GACTAATTAG AATATTGTAT CCTGTAACAG	12900
AGGCAACTTT GTCAGTAAAA TTCCGTAAAA TAATGGACTT TATTAAGTTT ACATCTGCTT	12960
GATTATTTAA AATGATAAAA ATCGGGATAG CAGGTAGTGA GGAAAAGATG GTTCTGTCA	13020
AGTAGAGTGA GAAAAGGTAC AGCCGATGCT GGTGATAAC TCCTTCAATC TTCTGCTCAG	13080
TCATCCACTC TTGAACAATT GCTTTCGAAA TATGATACAG TGGCTTGTCG CTTTCAATCC	13140
CATAATGTTT GTAATAATTA TAATAGGGAA CTAGATTTTG TAAACCAAAC AAAACGTTT	13200
TTGTTAAGAA AGTCAGTGCT GTTAAAAAG AAAGAGAATT CGAAATGTCA TTTCTTAAGA	13260
TATTTCTGAA CTTGGATAGT AGATGCTTTC CTCTGTATG CTGAAGAATC AGTTGAATAG	13320
TATGAGTCTT TTTTCTTGA TTCCATTTGT CTTGGAAAA CGAAGAATTA GCAGAACAAT	13380
AAACCAAAA GATATAATCC AGTTCTTCCT GAGTAAAAGT CATGTTGGCA TGTGGCTCTA	13440
AGTAAGTTTG GCAATGTTCC ATCAAAATCG GATACATAAA GAGGTTTTTT AATTTTTCAA	13500

486

ACTCTTTGGA CTCAGGGAAC TCAAGTGGAA ATTCCTCGACG TTTCCAAGTG AGTGCCACTA	13560
GTATGCTAAA ATGAACATAC TCGTCAGGTG TGATTTCCTAA CAGTTCATGA CTGAGTTGAG	13620
AATTAGACTG CACAATCATA TGTGTGACCC AATCCATACT TCCATCATTC AAATCATAAA	13680
TCTCAATACC AAAATGAAAC TGGAGGAGTG CAATTAAAAA ACGAATGCCA TATTTCAGGAC	13740
CAACTACTTG ATTTTTCACA AGGTCCAAAC CTACTGAACG TAGTAACAAG CCACACTTTT	13800
GTCGTACGCG GTAGCCTGTT GCGATGGAAA TATACTCTTT TTGTGTAAAT TCGTTAAAGC	13860
TTTGATTACC TTGTAGTAGA AAGAAGCGGA GTATTTTAA AATAGTTGAT TGGTTATAAA	13920
GCTGATGGAA GTAATAATTC GTTTGATGAG AATGGTGTTT GATTAATTGA ACTTGTGCG	13980
TATCTAAATT AAATGTCAAC TCTTCCTCGA ATGTTTCTTG TAATTCCTGC AAAATGCTTA	14040
GGAGACTTTT AGATTGTAAT GAAGTTAAAG TAGACAGTTC ATCTAGTTCA ATAGACCGAA	14100
TATCCAATAA TATATTTAAA ATGGTAATTT TATCTGTAAT TCTTTTTTCA ATGTATTTGT	14160
TTAGCATAGT TACCGAATCT TAGTTGCATA TAGATAATTT TAATTATTAT AATACAAAAG	14220
AAACTAATTG TCTTGTCAAA AAGTTGTGG AATTTCCGAC TTTATTGATA AAACAGCATG	14280
TAATAAAAGG CATTTTAAAG ATAGTAATGA GTATTGGTGG AGTTTTATGG CTTATTTTTT	14340
TTATTAGAAA ATATTTTTTT ATCAATAAT GTCGTTCTAT AAAAAATAT GTGATAAAAA	14400
TATCTATTGT GATGGAAGTT GTTTTAATTT ATACTAGGAT AGTTAATAGT AATACTATAC	14460
TATACTATAT TGTATACAAG TGTGTCATTG CCAGGTGAG AAGATAGCTA TAACGCACTT	14520
TTATACGCTT TTGCTACGTT TGTTAGTGAA CGGATTAACT CAGTGAGATA AATTTTATCA	14580
GAACATAAGT AATCCGTTTC TTCGTGTATA CAGATTGAAA GTACCTATGA ATCATAGAAG	14640
GATTAACTTG TTCTATGAAT AATGCTTAAC AGGGAGACAC ACATGAAAAA AGTAAGAAAG	14700
ATATTTTACA AGGCAGTTGC AGGACTGTGC TGTATATCTC AGTTGACAGC TTTTCTTTCG	14760
ATAGTTGCTT TAGCAGAAAC GCCTGAAACC AGTCCAGCGA TAGGAAAAGT AGTGATTAAC	14820
GAGACAGGCG AAGGAGGAGC GCTTCTAGGA GATGCCGTCT TTGAGTTGAA AAACAATACG	14880
GATGGCACAA CTGTTTCGCA AAGGACAGAG GCGCAAACAG GAGAAGCGAT ATTTTCAAAC	14940
ATAAAACCTG GGACATACAC CTTGACAGAA GCCCAACCTC CAGTTGGTTA TAAACCTCT	15000
ACTAAACAAT GGACTGTTGA AGTTGAGAAG AATGGTCGGA CGACTGTCCA AGGTGAACAG	15060
GTAGAAAATC GAGAAGAGGC TCTATCTGAC CAGTATCCAC AAACAGGGAC TTATCCAGAT	15120
GTTCAAACAC CTTATCAGAT TATTAAGGTA GATGGTTCGG AAAAAACGG ACAGCACAAAG	15180
GCGTTGAATC CGAATCCATA TGAACGTGTG ATTCAGAAG GTACACTTTC AAAGAGAATT	15240
TATCAAGTGA ATAATTTGGA TGATAACCAA TATGGAATCG AATTGACGGT TAGTGGGAAA	15300

487

ACAGTGTATG AACAAAAAGA TAAGTCTGTG CCGCTGGATG TCGTTATCTT GCTCGATAAC	15360
TCAAATAGTA TGAGTAACAT TCGAAACAAG AATGCTCGAC GTGCGGAAAG AGCTGGTGAG	15420
GCGACACGTT CTCTTATTGA TAAAATTACA TCTGATTACAG AAAATAGGGT AGCGCTTGTG	15480
ACTTATGCTT CCACTATCTT TGATGGGACC GAGTTTACAG TAGAAAAAGG GGTAGCAGAT	15540
AAAAACGGAA AGCGATTGAA TGATTCTCTT TTTTGAATT ATGATCAGAC GAGTTTACA	15600
ACCAATACCA AAGATTATAG TTATTTAAAG CTGACTAATG ATAAGAATGA CATGTAGAA	15660
TTAAAAATA AGGTACCTAC CGAGGCAGAA GACCATGATG GAAATAGATT GATGTACCAA	15720
TTCCGTGCCA CTTTTACTCA GAAAGCTTTG ATGAAGGCAG ATGAGATTTT GACACAACAA	15780
GCGAGACAAA ATAGTCAAAA AGTCATTTTC CATATTACGG ATGGTGTCCTCC AACTATGTCG	15840
TATCCGATTA ATTTTAATCA TGCTACGTTT GCTCCATCAT ATCAAAATCA ACTAAATGCA	15900
TTTTTTAGTA AATCTCCTAA TAAAGATGGA ATACTATTAA GTGATTTTAT TACGCAAGCA	15960
ACTAGTGGAG AACATACAAT TGTACGCGGA GATGGGCAAA GTTACCAGAT GTTTACAGAT	16020
AAGACAGTTT ATGAAAAAGG TGCTCCTGCA GCTTTCCAG TTAAACCTGA AAAATATTCT	16080
GAAATGAAGG CGGCTGGTTA TGCAGTTATA GCGATCCAA TTAATGGTGG ATATATTTGG	16140
CTTAATTGGA GAGAGAGTAT TCTGGCTTAT CCGTTTAATT CTAATACTGC TAAATTACC	16200
AATCATGGTG ACCCTACAAG ATGGTACTAT AACGGGAATA TTGCTCCTGA TGGGTATGAT	16260
GTCTTTACGG TAGGTATTGG TATTAACGGA GATCCTGGTA CGGATGAAGC AACGGCTACT	16320
AGTTTTATGC AAAGTATTTT TAGTAAACCT GAAAACTATA CCAATGTTAC TGACACGACA	16380
AAAATATTGG AACAGTTGAA TCGTTATTTC CACACCATCG TAACTGAAAA GAAATCAATT	16440
GAGAATGGTA CGATTACAGA TCCGATGGGT GAGTTAATTG ATTTGCAATT GGGCACAGAT	16500
GGAAGATTTG ATCCAGCAGA TTACACTTTA ACTGCAAACG ATGGTAGTCG CTTGGAGAAT	16560
GGACAAGCTG TAGGTGGTCC ACAAATGAT GGTGGTTTGT TAAAAATGC AAAAGTGCTC	16620
TATGATACGA CTGAGAAAAG GATTCGTGTA ACAGGTCTGT ACCTTGGAAC GGATGAAAA	16680
GTTACGTTGA CCTACAATGT TCGTTTGAAT GATGAGTTTG TAAGCAATAA ATTTTATGAT	16740
ACCAATGGTC GAACAACCTT ACATCCTAAG GAAGTAGAAC AGAACACAGT GCGCGACTTC	16800
CCGATTCCTA AGATTCGTGA TGTCCGGAAG TATCCAGAAA TCACAATTTT AAAAGAGAAA	16860
AAACTTGGTG ACATTGAGTT TATTAAGGTC AATAAAAATG ATAAAAACC ACTGAGAGGT	16920
GCGGTCTTTA GTCTTCAAAA ACAACATCCG GATTATCCAG ATATTTATGG AGCTATTGAT	16980
CAAAATGGCA CTTATCAAAA TGTGAGAACA GGTGAAGATG GTAAGTTGAC CTTTAAAAAT	17040

488

CTGTCAGATG	GGAAATATCG	ATTATTTGAA	AATTCTGAAC	CAGCTGGTTA	TAAACCGTT	17100
CAAAATAAGC	CTATCGTTGC	CTCCAAATA	GTAAATGGAG	AAGTCAGAGA	TGTGACTTCA	17160
ATCGTTCCAC	AAGATATACC	AGCGGGTTAC	GAGTTTACGA	ATGATAAGCA	CTATATTACC	17220
AATGAACCTA	TTCCTCCAAA	GAGAGAATAT	CCTCGAACTG	GTGGTATCGG	AATGTTGCCA	17280
TTCTATCTGA	TAGGTTGCAT	GATGATGGGA	GGAGTTCTAT	TATACACACG	GAAACATCCG	17340
TAAAGTGTAG	AAATGATAAT	ATCTATGTTC	TGAACGATAC	TTTTAAGAAG	TAGCACTCAA	17400
GAAGAGATTT	AAGTTTACTT	GGTGAAACCT	GTTTTATTTCG	TAAGTAAACT	ATCATTGAAA	17460
GGGGAGATGT	TTTCGAAAAC	TTGCACAGAA	AAAGGATTAT	TATTGTCATG	TGTAATTCAT	17520
TACATTGCTC	ACAGTTGATT	TTAAGAGATA	TGAATAAGGA	GAAATCATGA	AATCAATCAA	17580
CAAAATTTTA	ACAATGCTTG	CTGCCTTATT	ACTGACAGCG	AGTAGCCTGT	TTTCAGCTGC	17640
AACAGTTTTT	GCGGCTGGGA	CGACAACAAC	ATCTGTTACC	GTTTATAAAC	TATTGGCAAC	17700
AGATGGGGAT	ATGGATAAAA	TTGCAAAATGA	GTTAGAAACA	GGTAACTATG	CTGGTAATAA	17760
AGTGGGTGTT	CTACCTGCAA	ATGCAAAAGA	AATTGCCGGT	GTTATGTTTCG	TTTGGACAAA	17820
TACTAATAAT	GAAATTATTG	ATGAAAATGG	CCAAACTCTA	GGAGTGAATA	TTGATCCACA	17880
AACATTTAAA	CTCTCAGGGG	CAATGCCGGC	AAC TGCAATG	AAAAAATTAA	CAGAAGCTGA	17940
AGGAGCTAAA	TTTAACACGG	CAAATTTACC	AGCTGCTAAG	TATAAAATTT	ATGAAATPCA	18000
CAGTTTATCA	ACTTATGTCG	GTGAAGATGG	AGCAACCTTA	ACAGGTTCTA	AAGCAGTTCC	18060
AATGAAATT	GAATTACCAT	TGAACGATGT	TGTGGATGCG	CATGTGTATC	CAAAAAATAC	18120
AGAAGCAAAG	CCAAAAATTG	ATAAAGATTT	CAAAGGTAAA	GCAAATCCAG	ATACACCACG	18180
TGTAGATAAA	GATACACCTG	TGAACCAACA	AGTTGGAGAT	GTGTAGAGT	ACGAAATTGT	18240
TACAAAAATT	CCAGCACTTG	CTAATTATGC	AACAGCAAAC	TGGAGCGATA	GAATGACTGA	18300
AGGTTTGGCA	TTCAACAAAG	GTACAGTGAA	AGTAACTGTT	GATGATGTTG	CACTTGAAGC	18360
AGGTGATTAT	GCTCTAACAG	AAGTAGCAAC	TGGTTTTGAT	TTGAAATTAA	CAGATGCTGG	18420
TTTAGCTAAA	GTGAATGACC	AAAACGCTGA	AAAACTGTG	AAAATCACTT	ATTGGGCAAC	18480
ATTGAATGAC	AAAGCAATTG	TAGAAGTACC	AGAATCTAAT	GATGTAACAT	TAACTATGG	18540
TAATAATCCA	GATCACGGGA	ATACTCCAAA	GCCGAATAAG	CCAAATGAAA	ACGGCGATTT	18600
GACATTGACC	AAGACATGGG	TTGATGCTAC	AGGTGCACCA	ATTCCGGCTG	GAGCTGAAGC	18660
AACGTTTCGAT	TTGGTTAATG	CTCAGACTGG	TAAAGTTGTA	CAAACTGTAA	CTTTGACAAC	18720
AGACAAAAAT	ACAGTTACTG	TTAACGGATT	GGATAAAAAAT	ACAGAATATA	AATTCGTTGA	18780
ACGTAGTATA	AAAGGGTATT	CAGCAGATTA	TCAAGAAATC	ACTACAGCTG	GAGAAATTGC	18840

TGTCAGAAGAC TGGAAAGACG AAAATCCAAA ACCACTTGAT CCAACAGAGC CAAAAGTTGT 18900
TACATATGGT AAAAAGTTTG TCAAAGTTAA TGATAAAGAT AATCGTTTAG CTGGGGCAGA 18960
ATTTGTAATT GCAAATGCTG ATAATGCTGG TCAATATTTA GCACGTAAAG CAGATAAAGT 19020
GAGTCAAGAA GAGAAGCACT TGGTTGTTAC AACAAAGGAT GCTTTAGATA GAGCAGTTGC 19080
TGCTTATAAC GCTCTTACTG CACAACAACA AACTCAGCAA GAAAAAGAGA AAGTTGACAA 19140
AGCTCAAGCT GCTTATAATG CTGCTGTGAT TGCTGCCAAC AATGCATTG AATGGGTGGC 19200
AGATAAGGAC AATGAAAATG TTGTGAAATT AGTTTCTGAT GCACAAGGTC GCTTTGAAAT 19260
TACAGGCCTT CTGTCAGGTA CATATTACTT AGAAGAAACA AAACAGCCTG CTGGTTATGC 19320
ATTACTAACT AGCCGTCAGA AATTTGAAGT CACTGCAACT TCTTATTCAG CGACTGGACA 19380
AGGCATTGAG TATACTGCTG GTTCAGGTAA AGATGACGCT ACAAAGTAG TCAACAAAAA 19440
AATCACTATC CCACAAACGG GTGGTATTGG TACAATTATC TTTGCTGTAG CGGGGGCTGC 19500
GATTATGGGT ATTGCAGTGT ACGCATATGT TAAAAACAAC AAAGATGAGG ATCAACTTGC 19560
TTAAGTAAGA GAGAAAGGAG CCATTGATGA CAATGCAGAA AATGCAGAAA ATGATTAGTC 19620
GTATCTTCTT TGTTATGGCT CTGTGTTTTT CTCTTGATG GGGTGACAT GCAGTCCAAG 19680
CGCAAGAAGA TCACACGTTG GTCTTGCAAT TGGAGAACTA TCAGGAGGTG GTTAGTCAAT 19740
TGCCATCTCG TGATGGTCAT CGGTGCAAG TATGGAAGTT GGATGATTCG TATTCCTATG 19800
ATGATCGGGT GCAAATTGTA AGAGACTTGC ATTGCTGGGA TGAGAATAAA CTTTCTTCTT 19860
TCAAAAAGAC TTCGTTTGTG ATGACCTTCC TTGAGAATCA GATTGAAGTA TCTCATATTC 19920
CAAATGGTCT TTAATATGTT CGCTCTATTA TCCAGACGGA TGCGGTTTCT TATCCAGCTG 19980
AATTTCTTTT TGAAATGACA GATCAAACGG TAGAGCCTTT GGTCATTGTA GCGAAAAAAA 20040
CAGATACAAT GACAACAAAG GTGAAGCTGA TAAAGGTGGA TCAAGACCAC AATCGCTTGG 20100
AGGGTGTCGG CTTTAAATTG GTATCAGTAG CAAGAGATGT TTCTGAAAAA GAGGTTCCCT 20160
TGATTGGAGA ATACCGTTAC AGTTCTTCTG GTCAAGTAGG GAGAACTCTC TATACTGATA 20220
AAAATGGAGA GATTTTGTG ACAAATCTTC CTCTTGGGAA CTATCGTTTC AAGGAGGTGG 20280
AGCCACTGGC AGGCTATGCT GTTACGACGC TGGATACGGA TGTCCAGCTG GTAGATCATC 20340
AGCTGGTGAC GATTACGGTT GTCAATCAGA AATTACCACG TGGCAATGTT GACTTTATGA 20400
AGGTGGATGG TCGGACCAAT ACCTCTCTTC AAGGGGCAAT GTTCAAAGTC ATGAAAGAAG 20460
AAAGCGGACA CTATACTCCT GTTCTTCAA ATGGTAAGGA AGTAGTTGTA ACATCAGGGA 20520
AAGATGCTCG TTTCCGAGTG GAAGGTCTAG AGTATGGGAC ATACTATTTA TGGGAGCTCC 20580

490

AAGCTCCAAC TGGTTATGTT CAATTAACAT CGCCTGTTTC CTTTACAATC GGGAAAGATA	20640
CTCGTAAGGA ACTGGTAACA GTGGTAAAA ATAACAAGCG ACCACGGATT GATGTGCCAG	20700
ATACAGGGGA AGAAACCCCTT GTATATCTTG ATGCTGTGTT CCATTTTGTT GTTTGGTAGT	20760
GGTTATTGTC TTACGAAAA ACCAAATAAC TGATATTCAT TGTACATCAT TATGAATAGG	20820
ATAGCAGGCT GAAGGGAAGA CCAGAGTACT CTGAGGTGAT GTTAATCAGG AATCATGGTG	20880
ATGTGGCATG AATCATCAAT AACGGATATG AGGCTGGGCA GATTGTGCCA GCCTCATTGT	20940
GGGTATTGT TTGTAAAACG ATAGGACTGG TCTGGTAATC ATTTTA	20986

(2) INFORMATION FOR SEQ ID NO: 55:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21040 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CCCAGCAAAA AGCCATCCGA AGATGACTTT TTTGCTATTT AATTCTGTA TAAGTACTT	60
CCAAGCCACG CTTAACAGCT GGACGATTGG CAATTTTTC TGCCCATTTT ACTAGATTTT	120
GATAACTTGA GGCATCCAAG AATTTTGAG AACCTTGTA AAGATTCCT TGAAGTAACT	180
GTCCATACCA AGACCAGATA GCAATATCTG CAATCGTATA GTCATTGCCT GCAATATAAG	240
GTTTCTGAGC CAATTCCTTA TCCAATAAAT CCAACTGGCG TTCACTTCC ATCGTAAAAC	300
GGTTAATAGG ATATTCCAAT TTTTCAGGAG CATAATGAA GAAATGTCCA AATCCCCAC	360
CTAGAAAAGG TGCTGCACCT GCTTGCCAGA ATAGCCAATT CAAAACCTCT ACCTTTTCCA	420
CAGGATTACT TGGTAAAAAG GCTCCAAAT TCTCAGCAAG GTAAAGAAGA ATATGAGCAG	480
ACTCAAAGAC TCTTACGTTT TCAGTACCTG ACTGGTCCAA TAAGGCTGGA ATCTTGGAAT	540
TTGGATTGAG CTTACAAAAG TCTGATCCGA ATTGATCCCC ATCCATGATA GCAATCTTAT	600
ACAAGTCGTA AGCCGCTTCC TTAAAACCAG CTTCTAGTAA TTCTTCCAAT AAGATAGTAA	660
CCTTCACACC ATTTGGTGTT CCCAGTGAAT AAAGCTGAAA AGCTTGTCT CCTTTTGGCA	720
AGTTTGTGTC GAAACGGGCA CCTGCTGTTG GTCTGTTTAG CCCCCTAAAA GTCCTTGAT	780
TACTAGCTTC ATCCTGCCAT ACGGTCGGTA ATTGATATGC TGACATCCGA AACCTCCCTT	840
AAATCGCATT CTGTCAAAA CCGAGTTTGC GTTGAATAAA CTTAACGATT TCGACGATGA	900
TAATCATTGA GAGCTTCCA GCCATAACAA TTCCCCATTG TGACAAGTCT AGTTTGGTTA	960
CGTGGAAGAT TCCTTCAAGC GGTCTACAA CGATTGTTGC CATGAGAAGG ATAAAGGATA	1020

491

CCAAGATGGA CCAGTTAAAG GTCTTAGACT TGAATGGGCC AACTGTCAAG ATGGATTGGT	1080
AGACAGACTT GACATTGTAG GCATGGAAGA GCTGAATCAA ACCAAGGGTT GCAAAGGCCA	1140
TCGTTAGGGC ATCTGCATGA ATAGCATGAT TGTCACCCAC ATGAACTGGG TAAGCAATCG	1200
CAAGGCCATA AACACTCATA ACAAGAGCTG CTTGGAGTAC ACCTTGATAA ATGATAGAAC	1260
TCAAAACACC ACCTGAGAAG AAGCTTGCCT TGCCTCCACG TGCTTTATGA TTCATGACAC	1320
CAGGTTCCGC AGGTTCAACA CCAAGAGCGA TAGCTGGGAA GGTATCCGTT ACCAAGTTGA	1380
TCCACAAAAG ATGAACCGGC TGTAAGACAT CCCAACCAA CAAGGTTGAT AGGAAGATGG	1440
TTAATACTTC AGCAGTATTA GCAGAAAGTA GGTACTGAAT AGTCTTTTGA ATGTTTGAGA	1500
AGACCTTACG TCCTTCTTCC ACTGCGACGA TAATAGTCGC AAAGTTATCA TCTGCAAGAA	1560
TCATATCAGA AGCCCCCTTA GAAACCTCTG TACCAGTGAT TCCCATACCG ATACCGATAT	1620
CGGCTGTTTT CAGAGCTGGC GCGTCATTGA CACCGTCACC TGTCATGGCA ACGACTTTAC	1680
CTTGTTTTTG CCAAGCCTTG ACGATACGAA CCTTGTTGTC TGGAGACACA CGGGCATAAA	1740
CAGAGTATTG ACCAACGACT TTTTCAAATT CTTCATCTGA CAGTTCATTG AGTTCAGCAC	1800
CAGTTAAAC GTGACCTTCT GTATCGTTTG CGTCAATGAT TCCCAAACGT TTGGCAATGG	1860
CTTCCGCTGT GTCTTGGTGG TCACCTGTAA TCATAATTGG ACGGATTCCC GCTTCCTTAG	1920
CCACACGAAC AGCCTCAGCG GCTTCAGGAC GTTCAGGGTC AATCATCCCA ATCAAACCAG	1980
TAAAAATTAA ATCATTTTCA AGCTCTTCAG AAGTGAGATT TTCTGGAATA CTATCGATAA	2040
TCTTATAAGC ACCTGCAAGG ACACGCAAGG CTTGATGAGC CATTTCAGAA TTGTTTGTAC	2100
GAATGAGATT TGTAACCTTC TCATCAATCG GAGCAATATC CCCAGCCTTA TCACGAAGAA	2160
GACAACGTTT TAAGAGTTGG TCTGGCGCAC CCTTGACTGC TACAAGGAAA CGACCATCTG	2220
GCAATGGGTG AACTGTTGAC ATGAGCTTAC GGTGAGATC AAATGGCAAT TCAGCTACAC	2280
GAGGATATTT CTCTAAGAAA CCTTTGACAT CATAGCCCTT GTCCAAGGCA TATTGGATAA	2340
AGGCTGTTTC GGTTGGGTCA CCAATCAAGT TACCTTCCAC ATCGATTTTC GTATCATTTG	2400
CCAAGACAAC TGAACGAAGT AGTGGCATT TCAAGACCTAG TTCAATATCA TCAGCTGAGT	2460
CATGTAGAAC CGCATCGTAG AAGACTTTTT CGACTGTCAT CTTGTTTCATA GTCAGCGTAC	2520
CAGTCTTATC AGAAGCGATG ATTTCAAGTTG AACCAAGTGT TTCAACTGCT GGCAACTTAC	2580
GAACGATGGA ATGTCGTTTG GCCAAAACCTT GAGTACCAAG AGAAAGAACG ATGGTAACGA	2640
TAGCAGGAAG TCCTTCTGGA ATGGCTGCAA CGGCAAGGGC AACAGAAGTC AACAACTCAC	2700
CAAGTGGAAT TTTCCCTTGA ATGAAGACAC CCACTACAAA AGTAACAAGG GCAATGACCA	2760

492					
AGATAGCATA	GGTCAAGACC	TTAGAAAGGT	TGTTCAAATT	TTGTTTGAGT	GGTGTATCAG
2820					
TCTCATCCGC	ATCTTGAAGC	ATACCAGCAA	TATGACCAAC	TTCAGTGTAC	ATACCTGTAT
2880					
TGACAACAAC	ACCCATCCCA	CGACCATAGG	TTACGTTTGA	GTTTTGGAAG	GCCATGTTGA
2940					
CACGGTCACC	AATACCAGCA	TCTGTCGCAA	GCTCGACTGA	CAAGTCTTTT	TCGACTGGTA
3000					
CAGATTCAAC	TGTCAAGGCT	GCTTCTTCAA	TTTAAAGAGA	GTTGGCTTCT	ATCAAACGTA
3060					
GGTCCGCTGG	TACCACGTCA	CCTGCTTCAA	GGGCAACGAT	ATCGCCTGGT	ACCAATTCTT
3120					
TAGAGTCAAT	CTCTGCCATG	TGTCCATCAC	GAAGAACGCG	GGCAACTGGA	CTAGACATGG
3180					
ATTTGAGGGC	TTCAATAGCT	TCTTCAGCTT	TTCTTCTTGT	GTAAACACCA	AAGGCACCGT
3240					
TGATGATAAC	CACAGCTAGG	ATGATAATGG	CATCTGCGAT	ATCTTCCCCA	CCAGAAGTCA
3300					
CGACTGACAA	GATTGctGCC	GCAACTAGGA	TGATAATCAT	CAAACTCCTTA	AATTGCTCGA
3360					
TGAATTTGAC	CAAGATTGAT	CGTTTCTCGC	CTTCTTCGAG	TTTATTTGTG	CCAAATTCGG
3420					
CAAGGCGCTT	TTCCGCTCA	CTTGATGACA	AACCTTGCTC	GGTCGCATCC	ACAGCCTGCA
3480					
AGACCTCTTC	AGGGCTCTGA	GTATAAAACG	CTTGGCGTTT	TTGTTCTTTT	GACATGTGTC
3540					
TCCTCCTTGA	CATTGTGTGC	AAAACAGACT	CTCTTTCTGT	CATAGCTTTT	CACGACAAAC
3600					
AAAAAGAAAC	CTGTTAATCA	TAACAAGTCT	CGCTGTTTAA	GATAGGGCCG	GAAAGCATAC
3660					
TTTTTCAGCAT	AAAATTCGGA	ATGACGACAC	TATCACAGGT	TTCTGCCAGC	TACTCCCTTG
3720					
AGTAGTACCA	TTATACCAAA	TTTTGGGGAG	TTTTCAAAGA	GTAAAACTG	CCTTATTTGA
3780					
ATTTTTCCTT	GAAAACCACT	ATAATGGTAG	AATGCTATGT	GAAGTGAAG	GAAGTTGAAT
3840					
GAAGCAATCT	ATCTCAAATC	TCAAGTTAGC	TGAGCGTGGA	GCCATTATCA	GTATTTTCGAC
3900					
CTATTTGATC	TTGTCTGCAG	CCAAATTAGC	AGCTGGTCAT	CTCCCTCATT	CATCCAGTTT
3960					
GGTGGCCGAT	GGTTTTAATA	ACGTATCGGA	CATCATTTGA	AATGTGGCCC	TCTTAATCGG
4020					
GATTTCGGATG	GCGCGCCACC	TGCAGACCGT	GACCACCGTT	TTGGTCATTG	GAAGATTGAA
4080					
GATTTGGCAA	GCTTGATCAC	TTCTATCATC	ATGTTCTATG	TCGGTTTCGA	TGTTCTAAGA
4140					
GATACCATTC	AAAAGATTCT	CAGTCGGGAA	GAAACGGTCA	TTGATCCTCT	TGGTGCAACT
4200					
CTAGGAATCA	TTTCTGCAGC	GATTATGTTT	GTGGTCTATC	TCTACAATAC	TCGCCTCAGT
4260					
AAGAAATCCA	ACTCCAATGC	GCTGAAGGCA	GCTGCTAAGG	ACAATCTTTC	TGACGCTGTT
4320					
ACCTCACTTG	GAACCGCCAT	TGCCATCCTA	GCTAGTAGTT	TCAATTATCC	GATTGTGGAT
4380					
AAACTGGTTG	CTATCATCAT	CACTTTCTTT	ATCTTGAAGA	CTGCCTATGA	TATCTTCATC
4440					
GAGTCTTCCT	TTAGTCTTTC	AGATGGCTTT	GACGACCGCC	TGCTCGAGGA	CTACCAAAAG
4500					
GCTATCATGG	AAATTCCCAA	AATCAGCAAG	GTCAAATCGC	AAAGAGGTCG	CACCTACGGT
4560					

493

AGCAACATCT	ACCTGGATAT	TACACTAGAG	ATGAATCCTG	ACTTGTCTGT	TTTTGAAAGC	4620
CATGAAATCG	CGGATCAGGT	CGAGTCTATG	CTGGAGGAGC	GTTTGGCGT	CTTTGATACC	4680
GATGTCCATA	TCGAACCAGC	ACCTATCCCT	GAGGATGAAA	TTTGTAGACAA	TGTCTATAAA	4740
AAATTGCTTA	TGCGTGAACA	ATTGATTGAC	CAAGGAAACC	AACTAGAAGA	ACTCTTGACT	4800
GATGATTTTG	TCTATATTCG	CCAAGATGGA	GAGCAGATGG	ATAAAGAGGC	TTATAAGACC	4860
AAAAAAGAGT	TAAATTCTGC	TATCAAGGAC	ATTCAAATTA	CTTCCATCAG	TCAAAAAACC	4920
AAACTCATCT	GCTATGAGTT	AGATGGTATC	ATCCATACCA	GTATCTGGCG	TCGCCACGAA	4980
ACCTGGCAAA	ATATCTTTCA	TCAAGAAACC	AAAAAGAAT	AGAGAAATCC	TTTCATGAGA	5040
CGGGATTTTT	CTATCTTTT	ATACTCAATA	AAAATCAAAG	TGCAAATTAG	GAAGCCGGTC	5100
ACAGGCTGTA	CTTGAGTCGG	CAATGTGAAG	CCGACATAGT	TTGCACTTTG	ATTTTCGAAT	5160
AGTCTTAACT	ATCAAATFCA	CTGAGATACT	CATAGCGTTC	GTATTTTTC	AGGAGTGCTT	5220
CATTTTCTC	ATCCAATCT	TTTTGGAGAG	TAGCCAGCTT	ACCAAAGTCA	GAGCCGTTAG	5280
CCTGCATTT	CTCTCAATA	GCAGCGATAC	GTTTTTCCAA	GGTTTCAATA	TCACCTTCAA	5340
TACTTGCCCA	CTCCTGCTTT	TCTTGGTAGG	TCATGCGTTT	CTTGTCTTCT	CGAACCTTGA	5400
CCACTTTTTC	CTTTCGCGC	TTTTGCACCT	GATMGCCAT	ATCTGTTTCA	AAAGCTTTTT	5460
CATCAAGATA	GTCGGTGTA	TGACCAAAGA	AAGGACGAAT	CTTGCCATCC	TCAAAAGCGA	5520
GAATCTGGT	CGCTACCTTA	TCCAAGAAAT	AGCGGTCGTG	ACTGACTGTT	AAAACGGGAC	5580
CTGCAAAACC	TTGCAAGAAA	TTCTCTAAGA	CTGTCAAAGT	TGCAATATCT	AGGTCATTGG	5640
TTGGCTCGTC	TAAAAGAAGA	ACATTGGGTT	TTTCCAAAAG	CAGTTTGAGG	AGATAAAGAC	5700
GTTTTTCTC	ACCCCTGAC	AATTTCTCAA	TCAAAGTCCC	ATGCGTCGAA	CGTGGGAAGA	5760
GGAATTGCTC	CAGCAACTCA	GCGATGGAAG	TCGTAGAACC	ACCACTGGTC	TTGACCTCCT	5820
CTGCCACTTC	CTGCAGGTAA	TTGATCACAC	GCTTGCTTTC	ATCCAAACCC	TCAATTTGTT	5880
GAGAGAAATA	GGCGATGCGA	ACAGTTTCCC	CAATCACAAC	TTGTCTGCT	GTCGGCTCAA	5940
GACTTCCTGC	AATCAGGTTA	AGTAGGGTTG	ATTTTCCAAC	ACCATTGTCC	CCAACAATTC	6000
CAATACGGTC	TTTAGCCTGA	ACTAAGAGAT	TAAAATTTTG	CAAAATGGGC	TTATTTTCAT	6060
AGGCAAAGGA	AACATCCTGA	AACTCGATGA	CTTCTTCCC	AATCCGACTG	GTTTCAAAGT	6120
TCATAGTCAA	GTCTGTCTCA	GCACTACTGC	CTGAAACTTC	CTTTTTCAGA	TCATGGAAAC	6180
GATTGATACG	AGCTTGTTGC	TTGGTCGCAC	GCGCCTGCGG	TTGTCTGCGC	ATCCAGGCCA	6240
ATCTTGTGTT	GTAGAGTTGT	TCTTTTGT	GAAGAAGAGC	CGCGTCGCGC	TCATCCTGTT	6300

494

CCGCCTTTAG GCGAACATAG TCCTGGTAAT TTCCCTGGTA CTCGGTCAAG CCTGCACGAT	6360
CCAACTCGAA AATCCGTGTT GACAAAGCGT CTAAGAAATA ACGATCGTGA GTGATAAAAA	6420
GGACGGTCTT CTTAGAATTT TTCAAAAAGA GGGTCAGCCA CTCAATAATC GCAATATCCA	6480
GATGGTTGGT CGGCTCATCC AAAAGCAAGA GGTCGTGGTT GCCAAGTAAG ACTTGTGCCA	6540
ACTGTACCCG TCTTCTCAGA CCACCTGACA ATTCCCCAAC AGGAGTAGAT AAGTCTTGAA	6600
TGCCCAATTT GCTAAGAACG GTCTTGACCT GACTTTCGAT TTCCCAAGCT TGGAGAGAGT	6660
CCATCTCTGC CATGACACGT TCCAAACGCG CTGCTTGTC CTCACTATAG TCGAGCATAA	6720
TCAATTCATA CTCACGAATG AGCTGGATTT CCTTGAGTTC ACTAGATAGA ACCGTATCCA	6780
AAACTGTCTT TCTATCATCA AAATCAGGAT CTGAGTCAA GTAACCAATC TGGTAATCAT	6840
TTTTAGCTGA AAAAGGACTG ACATCCCCAT CAAATCCAGA AACACCAGAA AGGACGTCCA	6900
AAAGGGTGGT CTTGCCAGTC CCATTGACAC CGATTAAACC AATTCTGTCT AAGTCATGGA	6960
TAATAAAGGA AATATCCCTA AAAACGGTCT TGTCACCAAC GGATTTACTT AGTTTTCAA	7020
CGATAAAATC ACTCATTTTT TCTCCCTCAG GTAAGCATGG ATGGCTTCAC GATTATTCTC	7080
CAATCTCCA TCGACAATGG CAACTCAAT CTCTGTTAAA ATCTCTCCA AGTCTGGGCC	7140
TGGCTGATAG CCATATTCCT TGATCAAAAT ACCGCCATTA ATCTGAATCT CTTTCTTGTC	7200
ATGGATAGTC AAGCTTTGGT ATTTTCTGT GATGGCTTGT GGGTTGACTT CTTTCTCTG	7260
AGCTTGACGA AGATTTTCAG CCTGTAAAG CAAATCTATG TCAAAGCGAT AACATCTCG	7320
CTTGCTCAAT TCTCCATTTT CACGCAGAGC CAAAATAATC AGCAAATCCT GAACTTGCTT	7380
GGCAAAGTGG CGTGAGGTCT TCCAAGATTT CAAAATGAC TGCGCATTTT CAATCTCCAA	7440
AGCCCATAGT AAAGCCGCC AGGCTTGTTT AGAGGATTCA AAAGTAAAT CAGTCTCCAA	7500
ATCAAACAGT CTGTTGAGCT TGTCTGGCT AGATGCCATA TCAGGGAGAT AGTCATAAGC	7560
TTGACTCTCA ATCATGGAAG CCAAGCCCCT TCTCCAAAAT GGAGCCAGCA AGAGTTTATC	7620
AAACTCGACG AAGGTACGCT CTACAGAAAT TTTCTCCAA AGCGGCGTCA AGGTCTTCAT	7680
AGCTTTAAAT GTTTCTGGCT CAAGTGCAA ACCAAGACTA GCCTGAAAAC GGAAACCACG	7740
CATAATCCGT AAAGCATCTT CGTTGAAACG CTCACTAGCC ACTCCAAGT CTCGCAAGAC	7800
TTGCTTTTCC AAATCTTCTA AACCATGGAA CAAGTCAACG ATTTCTCCTG TCTCATCCAA	7860
GGCAAAGGCG TTGACTGTGA AATCACGGCG TTTGAGGTCT TCTTCTAGCG ATCGTACAAA	7920
GGAAACCGCA CTGGGTCTGC GATAGTCCAC ATAGACATCC TCTGTCCGAA AGGTGTTC	7980
CTCATACTCC TCATCCCCAT CTAAGACCAA GACGGTTCCA TGCTCGATTC CGATATCGGC	8040
TGTTGCGGGA AAAATCTGCT TGGTCTCTTC TGGATAAGAA GACGTCGCAA TATCCACATC	8100

495

GTGGATAGGG CTATGGAGAA GGGCATCTCG AACAGAGCCC CCAACAAAAT AAGCCTCAAA	8160
GCCTGCTTCT TTAATTTTTT CTAATACTGG TAAAGCCTTC TGAAATTCAG AAGGCATTG	8220
CGTTAATCTC ATAATAAGTG TTCTAATCCA TAGACAAGCT CATGACGCTT GACAACCTCT	8280
TTAATTCCCA AATTGACTCC TGTCATGAAG GAGATGCGAT CATAGGAGTC ATGACGGAGG	8340
GTCAACCCCT CTCCCTGATT GCCAAAGATG ACTTCCTGAT GAGCTACCAA GCCTGGCAAA	8400
CGAACTGAGT GGATGCGCAT ACCATCAAAG TCAGCACCAC GAGCACCAGC AATCAGCTCT	8460
TCCTCATCTG CTGCACCTTG CTGAATTGAC TCTCGAACCT CTGCCATCAA CTGAGCTGTT	8520
TTAATGGCTG TTCCACTCGG AGCATCCTTT TTCTTGTCAT GATGGAGCTC AATAATCTCC	8580
ACATTTGGGA AATATTTGGC AGCCTGCGTC GCAAATTGCA TGAGTAAGAC AGCACCACAAG	8640
GCAAAGTTAG GGGCAATCAG GCCACCCAAG TCTTGGGCAC GAGAAAATTC TTTTAGCTCT	8700
GCAATTTCTT CACTCGTGAA ACCAGTCGTT CCAACTACTG GAGCAAAGCC ATTTTCAAGA	8760
GCAAAACGTG TATTTTCGTA GGCAACAGCT GGAGTAGTAA AATCTACCCA GACATCCGCT	8820
TCAAAACCAG CTAATCAGC CTTATCCTTG AAAACAGGAA TACCCTGCCA TTCTGACTCA	8880
GACTCAAAAG GATCCAAAAC TGCCACCAAG TCCAAGTCTG GATCAGTCAA TACCATCTGA	8940
CAAGCAGCCT GGCCCATCTT TCCCTTAAAA CCGGCAATAA TTACTCGAAT ACTCATCTCT	9000
ACTCCTGTCT AAGATACAAA GTCCGTAAGA ACACAAAGTG AAAATAGGAA TTCCAATCAA	9060
GAAGTGCTA CTTCTTGGA GAACTATCTT TTTCACACAG GGTTCAGGC GTGTCAATT	9120
ATCAAGATAC AAAGGACCTT AGCTGCCTCT GAAAAATAGG GAATGGCACT GACTTTCCAC	9180
GAAAGGCAAG ACAGGCATCT TTTTCAAGA GGCAGGTAGT CCGTGTCAA TTTCTAAGAT	9240
ACAAGGCATC TTAAGTAGCC TAGAAGCGCC AACTAAATCA CTGGAATATA ACCCAGAGCA	9300
ATACTTCTTG CTCCTAGGTG CGTTCCAATG AACTACCAA ATGTAGCAAG TGAAACATCC	9360
GAACCAAGC CAAAATCAAG CAAGTGCTGA CGCAATTTCT CAGCCTTTTC AGGAGCATTC	9420
CCATGAATGA CAATGACCCG GTATTGACCT GAAGCCGTTG TTTCCTTGAT AATTTCAATT	9480
AAGCGCTTGG TGGCCTTCTT TTCAGTACGA ACTTTTTCGT AAACCTCAAT CACACCTTGA	9540
TCGTTAAAT AAAGGATTGG CTTAATGCTA AGCAAATTGC CCAAATGGC AGCCCCATTT	9600
GAAAGGCGTC CACCTTTTAC CAAATGATCC AAGTCATCTA CCATGATAAA GGCTGACGTA	9660
CGGCTGATTT GAATGGCTAG CTTATCCTGA ATGCTGGCAA AATCATCGCC CTGATCACGC	9720
CAATTAAAGA CGCTTTCAAC CATGATGCCT AGGGGAGCAC TTGTAATCAA AGTGTCTGGG	9780
AAAGCAATGG TTAAGCCCTC ATAGTCATCG ACCATATACT GGATATTTTG GTAAAAACCT	9840

496

GAAATTCCAG	AAGATAGGAA	AAGCCCCAAG	GCATGTGTAT	AGCCTTGTTT	TTTGAGCGAA	9900
GTTAAGATCT	CATCTAAGTT	GGCAATACTT	GGTTGACTGG	TCTTAGGCAA	TTCAGAAGCC	9960
TGAGCCATTT	TTTGGTAAAA	TTCCTCAGCA	GACAGATTGA	TGCCCTTCGAC	ATATTCTCTCA	10020
CCATCAATAT	TGACAGGAAT	ATCCAAGACA	AACAAGTCTT	CTCTTTGCAA	GATCTCTGCA	10080
CTGAGATAAG	CAGAGGAATC	TGTGAAAACA	GCTAATTTCA	TATTAGAACT	CCAAATTAAT	10140
TCCTGGTAAG	TCTAATGCAA	TTTCAGTCAC	TTGTAAGTTC	AAACGATTGA	GCATGTTCAA	10200
ACATGGACGA	GCCAAGGTTT	CCACCTCTTC	TTGGTTCAAT	TCATTGTTT	CATTGACAAT	10260
ACGGCCATCG	ATATGGTTTA	CTTGTGAGAT	TGTTCCACTA	ATGACAAACT	TATCAAATAC	10320
AATCATAAAG	CTCAAGATGA	CAATCAAGGA	AGTCACTTGA	TTTTCTTGGT	CATGTTGGAG	10380
CAATTGGAAA	TTACATCCA	CCTTGGTTTC	AGGAGCTCCA	TTTTCATTTT	CCCATTCAAA	10440
ATTACGCGCA	TCAAAATGAT	ACTGACTAAC	AAATCTTGT	TCACGTTTAA	GATTTCATGTC	10500
TTTCTCCATC	GGCTACAATA	TTATAAGCTA	TTGTACCATA	ATTTTTTATT	TTTCTCTAGT	10560
TTTCTAGGAT	TTAGTCAATC	CCAATTTTCA	CACGAACACT	ATCTGTGATG	GTATCAACAT	10620
AGTAGTTTAC	TTCTTCTGTT	GTAGGCGCTT	CTGCCATAAC	ACGCAAGAGG	GGTTCTGTTC	10680
CACTTGGACG	AACAAGGATA	CGGCCGTTCC	CCGCCATTTC	TTCTTCCATC	TTCTCGATGA	10740
TGGCCTTGAT	AGCTGGCACT	TCCATGGCCT	TTTCTTTCAT	GACGTTTTC	ACTCGGATAT	10800
TAATAATTTT	TTGTGGATAA	ATCGTTACTT	CTGCCGCCAA	CTCTGATAAG	CTCTTACCAG	10860
TTTCTTTCAT	GATTTTAGTC	AATTGAACTG	CTGATAATTG	ACCATCACCT	GTGGTATTGT	10920
AATCCATCAA	GATAACGTGA	CCAGACTGTT	CACCACCAAG	GTTGTAGCCT	GATTTTCTCA	10980
TTTCTTCAAC	AACGTAGCGG	TCACCAACTG	CAGTAACTGC	CTTGTTAATA	CCTTCGCGAT	11040
TCAAGGCCTT	GTGAAACCA	AGGTTAGACA	TAACAGTTGT	CACAATTGTA	TTTTGAGCCA	11100
ATTGTCCTTT	TTCAAGAAAG	TATTTTCCGA	TGATGTACAT	AATCTTGTCA	CCATCAACGA	11160
TGTCACCATT	CTCATCAACA	GCAATCAAGC	GGTCACTGTC	TCCATCAAAG	GCCAAACCAA	11220
TAGCTGACCC	ACTTTCTTTG	ACCACCTCTT	GAAGGGCTTC	TGGATGTGTT	GAACCAACAT	11280
TAAGGTTGAT	GTTAAGACCG	TCTGGTGTTC	CCCCGATAAC	CGTCAATTGG	GCACCAAGGT	11340
CTGCAAAGAT	TTGACGGGCA	CTGGTAGAAG	CTGCTCCATT	AGCTGTATCC	AAGGCAACCT	11400
TCATTCCATC	AAGAGGAGTT	CCAGTTGAAA	CAAGGTATCC	TTTCACTTCA	CGCAAGCTTC	11460
TGGATAATCT	ACCAAAATTC	CTAAGCCTTC	TGCACTTGGA	CGAGGAAGAG	TGTCTTCTTC	11520
AGCATCTAGC	AAGGCTTCAA	TTTCTGCTTC	TTTTTCATCA	TCTAGTTTGA	AGCCATCACC	11580
GCCAAAGAAC	TTGATTCCGT	TATCAAGGGC	TGGGTTGTGG	CTAGCAGAAA	TCATGACACC	11640

GGCACTTGCT CCTTCAGTTT CAACCAAGTA AGCTACTGCT GGTGTTGCAA GGACACCAAG	11700
TTTGATATACG TGAATCCCTA CTGAAAgGAG ACCTGCCACC AAGGCCGATT CCAACATTTT	11760
CCCTGAAATA CGTGTGTAC GTCCTACAAA GACTTTCGGC GCTTCCGTTT CATGTTGACT	11820
AAGAACATAG CCTCCAAAAC GTCCTAGTTT AAAGGCTAAT TCTGGTGTTA GTTCTAGGTT	11880
AGCTTCTCCA CGGACTCCAT CAGTCCCAA ATATTTACCC ATTGTTATAA AATCCTTTTC	11940
TATTTTAAAT TCGTTTTTGA ACTAGTTGCT TTCGTTGACG AAGATGTCTC CGATGAACTG	12000
CTTGTAATTG AATTGATGT GCTTGAACCT GGTGCTACTG GTTTGTAGT CACCTTCATT	12060
ATTGTATCAA ACGGAGTGAT AACTGCCGGT AAGACAACAC CATTGCGGTC GATTGCCTGC	12120
AAAGGTACTG AACCCTGTA ATTACCTGTT ATACGTTTCG TAGTTGGCAA AACAGCGATA	12180
ATCTTATCAA TTCTATCAA TGTCTCTTGG TCACTCGTAA TAGACACTTC TTTATCTGAC	12240
ACCATGACAT TTTCAATTTG TACCCGACTA TCAATTTGAC TAGGGTCAAT CTCGTGTACA	12300
ATCTTTACCT TATCCTTCTG AGCCTTCTTA CCAATCTTGA CTGTAATTTT TTGCGGAGTC	12360
GCCACAGCGG TCAGCCCAT GGGTAAATCT TCAATGCTCA AAGGAACCTC AATCGTTCCA	12420
ACACCGGCAT CTGTTAGGTC AGCAGTAACC TTGAATTTAC GTGTACTTTC TTGCATTTCA	12480
CTAGCTAGCG ATAGGCGATT TGCACCAGTC AAGACCACTG ATACTTCTGA AGCAAAACCG	12540
CTAATAAAAT ACTTATCACT ATTATAGCGT ATGTCAATAG GGACATTTGT TACTGTATTA	12600
GTATAGGTTT CCGTTTTTAC CTGCCTAGCA CTGGTACTGT TTTGAAAAT CGTCGCCGTA	12660
GCATAGACAA ATAAGACACA AGCAAAAAAG AGTGAGGATA TGATATATAA ACTATTTTTT	12720
TTTCATGTTT CATCCTCCTA GCAATCGTTC TTTAAACTA AGACCCACTT CCTCTTTTGG	12780
AAGTAAGATT TCACGTAATT CTGTTTCAAA TTCATCAAGT GTTAGGTTGT GCTTAAACCT	12840
TCCATTATAG GTTATCGAAA TTCCTCCCGT TTCCTCTGAT ACGACAAAAG TCAAGGCATC	12900
TGAGACTTCT GATAAACCGA TAGCCGCCCG GTGTCTGGTC CCAAATTCCT TGGAAATCCC	12960
TGTGTTTTTT GTCAAGGGCA GATAGGCAGA CGTCACAGCG ATACGTTCTT CTTTGATAAT	13020
CACCGCACCA TCATGTAGGG GAGTGTGGG AATAAAAATG TTAATGAGAA GTTCTGCAGA	13080
AATCTTAGCA TCCAAGGGAA TTCCTGTCGA AATATACTCC TGCAAGGTAC GTACACGCTG	13140
AATAGCAACC AAGGCCCGA TTTTACGAGG ACTCATGTAT TCAACAGACT TAACAAAGGC	13200
ACGAATCATC TGTTCCTCAG CACTAATAGG GGCATTGGAA AAGAAATCTG TCGCTCTTCC	13260
CAAACGTTCC AAACCAGTCC GAATCTCTGG AGAGAAGATA ACAACCGCCG CAATAACCCC	13320
ATAAGTAATA ATTTGATTGA TTAACCAAGA AATCGTAGTC AAACCAATCA TATTTGCAAG	13380

498

GATTTGAGCT AAAATAAACA CCAAACTCC ACGTACCAAA ATCATAATCT TGGTTCCTGC	13440
AATAGCTTTT GTAAAATGGT ATAAAATATA AGCAACAATC AAAATATCAA TCAGATTGAT	13500
AGCTATCGTC CATGGACTTG CAAACAACT GGTCCAATAT TGCAGATTGG ATAATTGTTG	13560
AAAATTTCATC CCTGATATCC TCCCTATCAA AACACTTTCG TCCTATTATA CCATTTTCTG	13620
GCATTTTTTT CCCTATCCTA GTCCATTTTA CATTGAACAA AAATATGATA AAATAAACTG	13680
ACTAAAAAAA ACAAAGGAGA AACTATGTCT CAACTCTATG ATATTACCAT TGTGGGTGGT	13740
GGTCCTGTCTG GGCTTTTTGC AGCCTTTTAT GCCCACCTAC GCCAAGCCAA GGTTCAAATC	13800
ATCGACTCTC TTCCCCAGCT AGGTGGACAA CCTGCTATTC TCTACCCTGA AAAGGAAATC	13860
CTAGACGTAC CAGGCTTCCC AAACCTGACT GGAGAAGAGT TGAATAACCG CTTGATTGAA	13920
CAGCTAAATG GATTTGATAC CCCTATTTCAT CTCAATGAAA CGGTTCTTGA GATTGACAAA	13980
CAAGAAGAAT TTGCCATCAC AACTTCTAAA GGAAGTCACC TGAATAAAC AGTTATCATC	14040
GCTATGGGTG GCGGTGCCTT CAAACCACGT CCGCTGGAAC TTGAAGGGGT TGAGGGCTAT	14100
GAAATATCC ACTACCACGT TTCTAACATT CAGCAATACG CTGGTAAGAA AGTGACGATT	14160
CTTGGTGGGG GAGACTCGGC TGTGGATTGG GCTTTGGCTT TTGAAAAAT CGCACCAACT	14220
ACCCTTGTTT ACCGCAGAGA TAATTTCCTG GCCTTGGAAC ACAGTGTTCA AGCCTTGCAA	14280
GAATCATCTG TAACCATCAA GACACCATTC GCCCCTAGCC AACTCCTTGG AAATGGAAAA	14340
ACACTTGATA AACTTGAAAT CACAAAAGTC AAATCTGATG AACTGAAAC CATTGACCTA	14400
GACCACCTCT TTGTCAACTA TGGTTTCAAA TCTTCTGTCG GTAACCTTAA AAATGGGGG	14460
CTCGACCTCA ACCGTCACAA GATTATCGTC AACAGCAAAC AGGAATCCAG CCAAGCAGGT	14520
ATCTATGCTA TCGGTGACTG CTGCTACTAT GACGGAAAAA TTGATCTGAT TGCACAGGC	14580
CTCGGAGAAG CTCCAACGTC TGTCAACAAC GCTATCAACT ACATTGACCC TGAACAAAAA	14640
GTACAACCAA AACACTCTAC TAGTTTATAA AAAAGAACCA CGAGTCACAT AGGATTCGTG	14700
GTTTTATAAT TCATCCGCTA TCTTATTGAT TTTTCTGAGT CTGTGATTGA CACCACTTTT	14760
GGTCAGAGGG GTGCTGAGAC TATCTGCTAA CTGCTGGATA GAGTAGTCTG GGTGCTGAAT	14820
CCTCAGTTGC GCCACTTCCT GCAAATCTAC TGGCAAATTT TCTAAGCCCA TGATATCTTT	14880
GATTTTACTG ATATTGTTAA TGGTCTTCAT GCTGGCAGAA ACTGTCCGAG CGATATTAGC	14940
TGCTCTCGCA TTATTAGCCC GATTGAGGTC GTTACGGGTT TCTCGCAAAA TCTTAACCCG	15000
CTCAAAATCA TCACGTGCCT GCATGGCTCC TATTACTATC AAGAAGTCCA TAATGTCTTC	15060
TGCTCGCTGG AGATAGGTCA CAGCCCCCTT CTTGCGCTCA AGCACCTTGG CATCCAGTAA	15120
AAACTGTTGG AGAAGGGAGG CAATTCCTTG CGCGTGGTCC AGATAAACAG AACTGATTTT	15180

499

CAACTGGTAC TTGCCTGACT CAGGGTCACG AATGCTCCCA TTTGCCAAGA AAGCGCCACA	15240
GAGATAGGCA CGACCTGCTT CCTCATCCGA TAAAATCGCC TCATCAATAC CTGTTTCCAG	15300
GCCAAAGAAA GAGTCTGCCA AGTGCAAATC ACTTAACAAA TCCTGCACCT TTTCATCTGT	15360
AAAAACGGTA TAGACGCGAT TCTTGCGAAG ATTGCTCCGT TGGTGGTGAC GAATTTTCAGA	15420
TTTGATTTC TAGAGATGGA GAAAGGACTC ATAGAGGTGA CGGGCCAGTT TGGCATTTC	15480
TGTCACAACT GACAAAGTCA AGCCCGAAGT CGAGAGACCG ATGCTACCAG ACAATTTTGAT	15540
AATGGCAGAT AATTCATGCC AGCTCAGATG GTGTTGGCCC AGGATTTCTT CTTTACTGC	15600
TACTGTGAAA CTCATTTTTC CACCTGTATA ATGCGCATCA ACTCGTCCAC AATCAAATCT	15660
CCATCGTGGA AGGCACCGCC ATTTTCCAGA CGAAGGAAGT TAGATGAAAT CACGCGCGAA	15720
ACTTGCTTAC AAAGACCTAC AAAATCGTGT TCCACTTGCA CTAAGTATTC ATCAAAACGG	15780
TTGGAATTCA TGTATTCCTG AGGCACCTTT TCAATATTCA CCAAGACAGT GTCGATAAAA	15840
GGGCGACCAA GGTGACGATG CAAGACTTCC ACGTGGTCGC TATCTGTAAA GTGTTCCGTC	15900
TCCCCACGTT GGGTCATGAT ATTGCAGACA TAGGCAATTT CTGCCTTGGT TTCCAAAAGA	15960
GCCCGCCCAA TTTCCTTAAT CACGATATTG GGCAAAATAG AGGTAAAGAG GGAACCTGGC	16020
CCTAGGACAA TCATGTCACT TTCAAGGATG GTCTGCACTA CTCGACGGCT GGCCAGAGGC	16080
GTATCATCGT TTAGGGCATT GGTACATAG ACATTGTCAA TTATGCCTCG ATGGTCTACA	16140
ATATGACTCT CTCAGCCAC TTCTGTCCCA TCCTGAAAGA CTGCATGAAG GGTCAAAGGA	16200
TGGTCACTGG AAGGATAAAT TTTCCTGTG GTATGGAAAA ATTTGCTCAA TAACTGCATG	16260
GCATTATAGG TTGAACCTG CATTTCTGAC AAGCCAGCAA TGATGAGATT TCCCAATGGA	16320
TGGCCAGCAA AGGCTCCGGC ATCCTCAGAG AACCGATACT GAAAGACCTT CTCATAAAAC	16380
TTAGGCATAT CCGACATGGC CACAAGGACA TTACGAAGAT CACCTGGCGG TGTCAACTGT	16440
TGCATATTTT TTCGGAGTTC ACCTGAAGAA CCACCATCAT CTGCCACCGT CACGATAGCT	16500
GCGATTTCCA CATCTTTTC CCGCAGACTT TTTAGAAATGA CGGGACTTCC AGTCCCTCCA	16560
CCAATCACCG TTATCTTTGG TTTTCTCATG AACGGTTTAC CGTTTCCITT CTGCGGTCTT	16620
TGTCGCGATG CCCTTCATTA ACAGACCAAT TCTTGATAA GTCCTGCGCC AAGCGTTTAG	16680
CAAATGCCAC ACTACGGTGT TGTCCACCCG TACATCCCAT GGCAATGGTC AAAACGGACT	16740
TACCTTCCTT TTGGTAACTT GGCAGAATCG GCTCAATCAA GGCCAATAAA TGTGATAAA	16800
AGTCTTCTGA CTCAGGATGG TTCATGACAT AATCATAAAC AGGTTTCATCC ACACCCGTTT	16860
GGTTTCTCAG TTCTGGTAAA TAATAGGGAT TTGGCAAGAA ACGGACATCA AAGACCAAGT	16920

500

CCGCATCAAT CGGGATTCCA TACTTAAATC CGAAAGACAT GACTTCGATA CGGAAAGACT	16980
GGGCTTGTTT TTGGTCTGAA AACTGCTCTG CAAGGGTTT GCGCAgCTCA CGTGGAGTGA	17040
GTTCAGTCGT ATCCACCACA TTTTGGCTCA TATTTTTCAG AGGTGCCAAG AGTTCACGTT	17100
CCAACTTGAT TCCATCTAAA ATACGACCGT CTGCTGCTAG TGGGTGACTC CGTCTGGTTT	17160
CCTTGTAACG AGCGACCAAT TCCTTATCAG CCGCATCCAA AAAGAGGATT TTGAAATCCA	17220
AACCATCTTG ATTTTCCAAC TCATCCAAAA CAGCTTGAAT CTCTGAAAAG AAAGAACGGC	17280
TACGCATATC CACTACCAAG GCCAACTTAG GATTGTCTTC CTTAATTTCA ACCAGCTGCA	17340
AAAACTTAGG CAAGAGAGCT GCGGCATAT TATCAATGGT GAAATAACCT AGATCCTCGA	17400
AGGACTGAAT GGCTACAGTT TTCCCTGCGC CACTCATCCC TGTCACAATC ACCAAGTGAA	17460
GTTGTTTCTT TGTCATCTTT TTCTCCTTAT ATCAAAAGAA GTTTGGCAAC ACCAACTTC	17520
AACTAGCTTA TCCAATCTCT GCGATGACTT CAATTTGAC TTTTACATCA CGAGGAAGAC	17580
GAGCTACCTC CACAGCTGAA CGAGCTGGGA ATTCTCTTT GAAGGCCGTT TGGTAAACCT	17640
CATTAAAAGG AACAAAGTCG TTCATATCGC TCAAGAAGCA AGTTGTTTGT ACAACATGGT	17700
CAAAGTCTGT TCCTGCTTCT GCCAAAATAG CACCGATGTT TTTCAAGACT TGCTCTGTCT	17760
GTTCTTGGAT ATTCTCTCCT ACAATTTCCC CAGTTTCAGG GGATAGGGGA ACTTGACCGC	17820
TAGCAAACAA AAGGTTGCCA ACGATTTTTC CTTGAACATA GGGTCCGATA GCCTTTGGGG	17880
CCTTATCTGT ATGAATTGTT TTTGCCATTT TCTTTTCCTC ACAATTTTTC TAAGATTGCA	17940
TCCCAGCCT CATCCATCCC TGCCTTACTG ACAGATGAAA AGAGGATGAA ATCGTCACTC	18000
GGGTCAAAGT TTAATTTCTT TTTGATTGCT GATTCATGCT TGTTCATTT ACCACGAGGA	18060
ATCTTGTCGG CCTTGGTCGC CACAATGATG ACTGGAATCT CATAATACTT GAGAAATCG	18120
TACATCTGCA CATCATCTGC TGACGGGTCA TGACGAAGGT CAACTAGACT GACAACCGCA	18180
CGGAGATTTT CCCGAGTCGT TAAGTACTCC TCAATCATGC ACCCCCACTT TTCACGTTCC	18240
TTTTTAGAAA CACGAGCATA GCCATAACCA GGCACATCCA CAAAGCGCAT CTTGTCATCA	18300
ATGTTAAAAA AGTTCAGGAG CTGGGTTTTC CCAGGTTTTC CTGATGTACG GGCGAGATTC	18360
TTACGGTCA ACATACTGTT GATAAAGCTG GATTACCAA CATTTGAACG CCCTGCTAGG	18420
GCAATCTCTG GCAGTTTATC CTGCGGATAG TGGGACTTAT TAGCTGCACT GAGCAAGATT	18480
TCAGCATGTT GTGTATTAAG TTCCATAGTC ACCTCTAGGC TGTTTCTAGG ATCGGTTTAT	18540
CCGTTCCATC TACAGTTTCT TTAGTGATGC GAACCAATTT CACATTTTCC TGACTCGGCA	18600
CCTCAACAT GACATCTAGC ATGGTTTCTT CGATGATGGA GCGAAGTCCA CGCGCCCTG	18660
TCTCCGTTT GATTGCTTTA TTAGCAATCT CTGAAGGGC TTCGTCGTCA AATTCCAAC	18720

501

CAACATCATC ATAAGAAAGC AAGGTTTGGT ATTGTTTCAC CAAGGCATTT CTTGGCTCTT	18780
TCAAGATGCG AACCAAGTCA TCAACGGTCA ATTGCTCAAG AGCCGCAAAA ACAGGCAAGC	18840
GTCCAATCAA CTCAGGGATA ATACCAAATT TTTGAATGTC TTCAGCGATG ATTTCTTGCA	18900
TGTATGAGCT GTTTTCGTCA ATCGCCTTAT TATTTTGACC AAATCCGATG ACTTTTTCAC	18960
CCAGACGTTG TTTGACAATT TCTTCAATAC CATCAAAAGC ACCACCCACG ATGAAGAGGA	19020
TATTTTTTGT ATCCACTTGA ATCATCTCTT GTTGTGGATG TTTGCGTCCA CCTTGAGGCG	19080
GTACGCTAGC AACAGTTCCC TCAATAATCT TGAGAAGGGC TTGTGCAACC CCTTCACCAG	19140
AAACATCACG TGTGATAGAC ACATTCTCAC TCTTCTTGGC AATCTTGTC AATTTCATCCA	19200
CATAGATAAT GCCACGCTCT GCACGTTCTGA TGTTAAAGTC AGCAACCTGC AAGAGTTTGA	19260
GGAGGATATT TTCCACATCC TCACCCACAT AACCAGCCTC CGTCAGAGCT GTCGCATCCG	19320
CAATAGCAAA AGGTACATTC AAGCTCTTAG CCAAGGTCTG GGCAAGGAAA GTTTTCCCTG	19380
AACCAGTTGG GCCAATCATC AAAATGTTTG ACTTCTGCAA ATCCACATCT TCTGACTCTT	19440
CGCGTGTATC GTGGAAATG ATGCGTTTGT AGTGGTTATA AACCGCCACT GCCAAGGCAC	19500
GCTTGGCAGC ATCTTGACCA ATTACATAGT GGTTCAGAT ATGGAGGAGT TCAATTGGTT	19560
TTGGCACCTC AGACAAGTCT GCCAAGACTT CCTCAACCAA TTCTTCTCGA ATGATTTCTT	19620
GAGCTAACTC CACGCATTCA TTACAAATAA AAGCATGTGTT GCCAGCAATT ATTTTGTGTA	19680
CTTCTTCTTG GTTTTGGCCA CAAAATGAGC AATAAACCAT CATATCATTT TTTCTATTTC	19740
TAGACATGAT TTCCTTCCAT TCTATACTGT CATTCTATCT AAAATAAGGT CATGTAAAAA	19800
GCATGAATAC TATTGACCAG ATTGGTAAAG GCATTTAACC AAAGGAGGAT AGAAAGCCCG	19860
TAACGCTTTT TACGAAAAGC TTGTGCTCCT GCCAGAAAGC AGATGAAACA CAGAAAAGCC	19920
GTGAATAGAC CAAATAAACT CCGTTCCATT AGACTTCCTT TCTCTTGCGG TATTGGATGG	19980
TAAAATCATA AGGATTCTTC TCATCTTTGG CGTAAAATTT GCTTGAAACT GTCTCAAAAA	20040
GAGACAAGTC AAGTTCTTCA GGGAAATAGG TATCTCCTTC CACCCGAGCA TGAATGTGAG	20100
TGACAATCAC TTCATCAAGG TAAGGTCAA AAGCCTGAAA AATTTGCTTC CCACCGATAA	20160
TGTAGAGATT CTTTTCTTGA GCCTGATACC AGTCAAGAAC AGACTGGACG TCCTGAAAAG	20220
TAGCAACCCC ATCTATCTTT TCTTCCGGAT TACGCGTCAA AATCAAGGTT TCCCCTTTTG	20280
GAAGCAAGCG ACGCCCCATC CCATCAAAGG TCACACGCCC CATCAAGATA GCATGATTCA	20340
GAGTTGTTTC TTAAAGTGC TGCAATTCTG CTGGCAAATG CCAAGGCAGA CGATTTCTCT	20400
TACCAATCAC ACCCTCTTCA TCCTGGGCCC AAATAGCTAC GATTTTCTTA GTCATGCTTC	20460

502

CATCCTTTTC ACTGATAGTA CTATTTTATC AAAAAACTCA AAAAAAGACT GGTTTGGAAT	20520
AGCTTACAAA ATAGAAAAAA TCTGTAAGAA ATTCCTTACA GATTATCTA TGTTCCTTA	20580
TTTCTTACAA ACCAGGTGCT TGTCCAAGTT CGGCTGCAAG CATCCAAATT GTTTATCTG	20640
TTTCAGTTTT AGCGCCTGCA AAGATACCGT TTGTCACATC GTCACCTTCT TCATCAGTGA	20700
CATCCAAACC TTTTGGAAA AGTTCTGACA AGTAACGGTA GATAACAAGA ACACGTCCA	20760
AGCTTCTTTC AACATTACGG TATTCACCAG CTTCTTCTTC GATTTCACTA TTTTGAAGGA	20820
ACTCTGTCAA TGTAAGAAAT GGGCTTCCAC CGAGTGTAAAT CAAGCGTTCA CTGATTTTCA	20880
CCAATTGACC GTCAAGAGCT TCCATGTACT CATCCATTTT TGGATGCCAT ACAAGGAAAC	20940
CACGACCATG CATATACCAG TGCACCTGGT GCAAAGCAAC GTGAGCTACA TACAAATCAG	21000
CAACAGCTTG GTTCAAGACT TCCTTTGTTT TTGCCAATGC	21040

(2) INFORMATION FOR SEQ ID NO: 56:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2387 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

ATTCTTAATA CGATTAAAG GCTTATTACT AAAAGAAAAT TTCAGTTAGA TGAACATAAC	60
TTGCTCGTCA AATCCCGATT TAACGAGATG TTTGGGGAAA ATAAATATT TGAAAGCATT	120
GATAACTTAT TTGATATTAT AGATGGTGAT AGGGGCAAAA ATTATCCTAA ATCAGATGAG	180
TTGTTTAGTG AGGAGTACTG TTTATTTTAA AATACAAAGA ATGTTACTAA AAACGGATT	240
TCATTCGATA CAAAGCAATT TATCACTAAA ACAAAGGATA AATTACTTCG AAAAGGCAAA	300
CTTGAGCGTT ATGATATAGT CTTGACAACA AGAGGTACTG TTGGAATGT AGCGTACTAC	360
GATGAATTAA TAAATATAA ACATTACGT ATAAATTCAG GTATGGTAAT ATTACGTCCC	420
AAGACACCAA ATCTAAATCA GAAATTTATT ATCCATGTTT TAAGGAATAA TAATTATAGT	480
CGAGTGATAT CAGGAAGTGC TCAGCCTCAG TTACCAATTA CAAAATTAAA AAAAATACTT	540
CTCCCCCTCC CCCCCTAGC CCTCCAAAT GAGTTCGCAG ACTTTGTAGT CCAGGTCGAC	600
AAATCACAAT TGGCAATCCA AAAATCTCTG GAAGAACTTG AAACCTTGAA GAAATCTCTG	660
ATGCAGGAGT ATTTTGGCTG ATATTCTGCC ATTGTAATTA CGGTAATGAT TTGTATAAT	720
ACTTCAAAGG AGGAAATCAG ATGGTAGTAA AAACAAGAAA ACAAGGAAAT TCAATCACCA	780
TTACGATTCC AAGTGAATTT AATATTCCAA GTGGTGTTAA ATACGAAGCG AAATGTGTAC	840

503

CAAGTGGTGA GATTATCTTT ACTCCTGAAG AATTGGGGCA GCAGGTTTCT TATGTATCTG	900
ATGATGCCCTT TGACTTAAAT TTAGATAAAA TATTTGACGA ATACGACGAT GTTTTCAAAG	960
CTTTGGTGGG AAAATGACAA TCTATTGAC AGAAAAGCAA ATTGAAAAAA TAAATGCTTT	1020
AGCAATTCAA CGGTATTCTC CAAATGAGAA AATTCAAACA GTTAGTCCTT CTGCCTTAAA	1080
TATGATTGTG AACTTACCAG AACAAATTGT CTTTGGGAAG CCTCTTTATC CAACAATTTT	1140
TGATAAAGCA ACGATACTAT TTGTCCAATT GATAAAGAAG CATGTTTTTG CTAATGCTAA	1200
TAAAAGAACT GCTTTCCTCG TTTTGGTCAA ATTTTACAA TTAAACGGCT ATCGTTTTTC	1260
TGTAACGGTA GAAGAAGCAG TAAAAATGTG TGTAACCATC GCAGTAGAAG CTTTAACTGA	1320
TGAAAAATG ACAAGCTACT CCAAATGGAT TTCTGAACAT TCTGTTAGAG AAAAGGTCAA	1380
AAAGTAACCT AGTATGCTGG ATTTGAATGA GCACAAGAAA ATAAATGAAC AGACAATATT	1440
AGAATTCTGT AATGCAGAAA CTGATATTGT CTCTTTTTAT TGATGAATAA GAAAGTGAGA	1500
AATTATGGAA TCAAAAGTTA CAATTATCAT GCAAGAAATG TTACCTCTTT TAAATAATGA	1560
ACAATTACTA GCGTTGAGAG AGAGTTTAGA ACATCATCTA GTAGACGGAA AAAAGCAGCA	1620
GAAGTATTCG AATAATAACC TGTGCAACT ATTTATTACC GCCAAGCAGG TAGAGGGCTG	1680
TAGCTCAAAA ACAATTCGTT ATTATCAGAG GACGATTGAA AACTTGTTTA ATGCTATTAA	1740
AGAGTCTGTG ACACAACCTA CAACAGATGA TTTAAGGAGT TATTTAGCAA ATTACCACTC	1800
TGAAAAGGAT TGTAGTAAGG CAAATTTAGA CAATATTAGG CGTATATTGT CTTCTTTTTT	1860
TGCTTGCTT GAGCAAGAGG ATATATCATT AAAATTCCCA TTCGACGGAT ACAGAAAATT	1920
AAGACTGAGC AAAATGTGAA GGAACTTAT ACTGATGAAC ATTTGGAAAT TATGCGTGAT	1980
AACTGTGAAA ATTTGAGAGA TTTGGCAATA ATAGACCTAC TAGCATCGAC AGGTATGCGT	2040
GTAGGGGAGC TTGTACAGTT GAATCGTTCA GATATTGATT TTGAAAACAG AGAGTGTGTT	2100
GTCTTTGGTA AAGGAAAGAA GGAGAGACCA GTATATTTTG ACGCTCGTAC GAAAATTCAT	2160
TTAAGAAATT ATCTTAACGA CAGAAAAGAT AGTCACCCTG CTCTTTTGT AACGCTAGTT	2220
GGAAAAGTCC AGAGGCTTGG AATTGCTGGT GTAGAGATTC GCTTAAGAAA GTTAGGAGAC	2280
AAACTCGGCA TACAAAAGGT TCACCCACAT AAGTTCAGAA GAACTTTAGC GACTAAGGCA	2340
ATTGATAAAG GTATGCCTAT CGAACAAGTC CAAAACTGC TAGGTCA	2387

(2) INFORMATION FOR SEQ ID NO: 57:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10669 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double

504

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

ATATTAAAGC GACTTTCTGT GCGCTAGGGA AAAATGTTCC TGGGAATGAG GACTTGGTGA	60
AGAGGATAAA ATCTGAAGGT CATGTTGTTG GAAACCATAG CTGGAGCCAT CCGATTCTCT	120
CGCAACTCTC TCTTGATGAA GCTAAAAAGC AGATTACTGA TACTGAGGAT GTGCTAACTA	180
AAGTGCTGGG TTCTAGTTCT AAATCATGTC GTCCACCTTA TGGTGCTATT ACAGATGATA	240
TTCGCAATAG CTGGATTGTT AGCTTTATCA TGTGGGATGT GGATAGTCTG GACTGGAAGA	300
GTAAAAATGA AGCATCTATT TTGACAGAAA TTCAGTATCA AGTAGCTAAT GGCTCTATCG	360
TTTTGATGCA TGATATTACAG AGTCCGACAG TCAATGCCTT GCCAAGGGTC ATTGAGTATT	420
TGAAAAATCA AGGTTATACC TTTGTGACCA TACCAGAGAT GCTCAATACT CGCCTAAAAG	480
CTCATGAGCT GTACTATAGT CGTGATGAAT AAGCAAGAAA AAATAGGTCT GTTAGATATT	540
TGACAGACTT ATTTTTTACA GAATATAGTA CTACTTAAAA AATGTTTTAT GCTATAATTG	600
ATGAATAAAA TAGAAGGAGA AGCATATGAA TACCTATCAA TTAAATAATG GAGTAGAAAT	660
TCCACTATTG GGATTGGA CTTTTAAGGC TAAGGATGGA GAAGAAGCCT ATCGTGCACT	720
GTTAGAAGCC TTGAAGGCTG GTTATCGTCA TATTGATACG GCGGCGATTT ATCAGAATGA	780
AGAAAGTGTG GGTCAAGCAA TCAAAGATAG CGGAGTTCCA CGTGAAGAAA TGTTGTAAC	840
TACCAAGCTT TGGAATAGTC AGCAAACCTA TGAGCAAACCT CGTCAAGCTT TGGAATAATC	900
TATAGAAAAA CTGGGCTTGG ATTATTTGGA TTTGTATTTG ATTCAATGGC CGAACCCTAA	960
ACCGCTCAGA GAAAAAGACG CATGGAAAAC TCGCAATGCG GAAGTTTGGG GAGCGATGGA	1020
AGACCTCTAT CAAGAAGGGA AAATCCGTGC TATCGGCGTT AGCAATTTTC TTCCCATCA	1080
TTTGGATGCC TTGCTTGAAA CTGCAACTAT CGTTCCTGCG GTCAATCAAG TTCGCTGGC	1140
GCCAGGTGTG TATCAAGATC AAGTCGTAGC TTAAGTCTCGT GAAAAGGGAA TTTTATTGGA	1200
AGCTTGGGGG CCTTTTGGAC AAGGAGAACT GTTTGATAGC AAGCAAGTCC AAGAAATAGC	1260
AGCAAATCAC GGAAAATCGG TTGCTCAGAT AGCCTTGGCC TGGAGCTTGG CAGAAGGATT	1320
TTTACCACCT CCAAAATCTG TCACAACCTC TCGTATTCAA GCTAATCTTG ATTGCTTTGG	1380
AATTGAACTG AGTCATGAGG AGAGAGAAAC CTAAAAACG ATTGCTGTTC AATCGGGTGC	1440
TCCACGAGTT GATGATGTGG ATTTCTAGAA AATCATAAAA AGAATGTAC ATTATTCTAA	1500
TTTTTGATAT AATAGTCAGC AGGAAAGAAA GTCTTATGGC GTCTTCAAG CGAGCTGGG	1560
ATAGTGGGAG CCAAGTAGGG CAAAATAAAG GGCTGGCGCT TTCTGTAGTA TTTTCAAAA	1620

505

CAATGAAGTA ATAAATTAGG GTGGAACCGC GTTCTGACG CCCCTAGGTT AAATCAACCT	1680
AGGATTGTCA GATGTGGTTC TTTTGCTTAT TCAGTCTATT GTGTGAAAGA AAGGAGAGCC	1740
GTGGACAACC TTTATCTTGT AAAAGACGAT AGTCAACTAG CTACATTTCG TGATTTTGTA	1800
GTAAGAAATA CTGAAAAGTT GAAAGATTAT CAATCTTTTT TAAAGAATGA ACTTGCAGTC	1860
TGTGATTTAC CGCAAGCTGT TATTTGGTCA GATTTTAATG CTGCTACACA GATTATTAGG	1920
GAAAGTGCTG TTCCAACCTA TACAAATAAT AGACGAGTGG TTATGACGCC TGATTTAGCT	1980
GTTTGGAAG AATTGTATTT GTATCAGTTG ATGGACTACG AGTGTTCGA GCAAACCTCA	2040
GCAATAGAAA GTCACATCA TTCTTTATCT GAAAATTTCC TCTTACAGAT TGTAGGACAT	2100
GAGTTAGCTC ATTGGTCGGA CATTTTTTAG ATGATTTTGA TGGTTATGAC TCTTATATCT	2160
GGTTCGAAGA GGGGATGGTT GAATATATTA GTCGCAAGTA TTTCTTGACA GAAGAGGAAT	2220
TTCAAGCGGA AAAAATTTGT AATCAATCTC TCGTAGAACT TTTTCAGAAG AAGTATAGTT	2280
GGCATTCAAT GAATGATTTT GGTTCCTCGA CTTATGATAA GAACTATGCA AGTATTTTTT	2340
ATGAATACTG GCGCAGCTTT TTGACAGTAG ATAAGTTGGT AGAAAATTTA GGTAGTGAC	2400
AAGCGGTCTT AGATTCTTAT CATTTATGGG CAAATACAGA AAAAATTTT CCCTTGTTAG	2460
ATTGGTTTGT TCAGCAGAAA TTAATTGAAA AAGAAATATA AAAACTAAAG GAGTAAACAA	2520
TGTCTAAGAA ATTAACATTT CACTGCATCA GTGGCAGAGA CCTCCTTACA GTCGGGCTGC	2580
TCCACGCTCA GCACTAGAGT GCCTGAGCTA GACGCAGTAC TAACTCGTCT TGCCTCGTAT	2640
GATCGACGAG GCAGACTCGT GTCGCAAGTA ATTATTTTTT ATTAAGGAGT ATTCAATGTC	2700
TAAGAAATTA ACATTTCACT GCGTCAGTGG CAGAAACCTC CTTACAGTCG GACTGCCCTA	2760
CGCTCAGCAC TAGAGTGCCT GAGCTAGACG CAGTACTAAC TCGTCTTGCC TCGTATAATC	2820
GACGAGGCAG ACTCGTGTCG CAAGAAATTA TTTTTTATTA AGGAGTATTC AATGTCTAAG	2880
AAATTAACAT TTCAAGAAAT TATTTTGACT TTGCAACAAT TTTGGAATGA CCAAGATTGT	2940
ATGCTTATGC AGGCTTATGA TAATGAAAAA GGTGCGGGGA CAATGAGTCC TTACACTTTC	3000
CTTCGTGCTA TCGGACCTGA GCCATGGAAT GCAGCTTATG TAGAGCCATC ACGTCGTCTT	3060
GCTGACGGTC GTTATGGGGA AAACCTAAC CGTCTCTACC AACACCACCA ATTCCAGGTG	3120
GTCATGAAGC CTTCTCCATC AAATATCCAA GAACTTTACC TTGAGTCTTT GGAAAAATTG	3180
GGAAATCAATC CTTTGGAGCA CGATATTCGT TTTGTTGAGG ACAACTGGGA AAACCCATCA	3240
ACTGGTTCAG CTGGTCTTGG TTGGGAAGTT TGGCTTGACG GAATGGAAAT CACTCAGTTC	3300
ACTTATTTCC AACAAAGTCG TGGATTGGCA ACTGGCCCTG TGAAGCGGA AGTTACCTAT	3360

506

GGTTTGGAGC GCTTGGCTTC TTACATTCAA GAAGTAGACT CTGTCTATGA TATCGAGTGG	3420
GCTGATGGTG TAAAATACGG AGAAATCTTT ATCCAGCCTG AGTATGAGCA CTCAAAATAT	3480
TCATTTGAAA TTTCGGACCA AGAAATGTTG CTTGAAAAC TTTGATAAGTT TGAAAAAGAA	3540
GCTGGTCGTG CATTAGAAGA AGGCTTGGTA CACCCTGCCT ATGACTATGT TCTCAAATGT	3600
TCACATACCT TTAATCTGCT TGACGCGCGT GGTGCCGTAT CTGTAACAGA GCGTGCAGGC	3660
TATATCGCTC GTATCCGTAA CTTGGCCCGT GTCGTAGCCA AAACCTTTGT CGCAGAACGC	3720
AAACGCCTAG GCTACCCACT TTTGGATGAA GAAACAAGAG CTAAACTCCT AGCAGAAGAC	3780
GCAGAATAAA GAGAGTGACA AATTACGAAA ATGGGCGAAC AGAGTGAGCC CTGAGCCAGT	3840
TGCCGCAGTG ATGAAGGTAT CCTTAGTGAA ACTAAGGATA CTAGGCAAAA TTGGAGACTT	3900
TTGGCTCCAA TTTAGCAAT GAAACAACGA AGTTGGTTGC TTGCGTGCCA ATCACATAAG	3960
GCAAACTGGA AAATAAAAAG ATACTTTTCG GAGAAAAAAC ATGACAAAAA ACTTATTAGT	4020
AGAACTCGGT CTGAAGAAT TACCAGCCTA TGTGTGTACG CCAAGTGAAA AACAACTAGG	4080
CGAAAAATG GCAGCCTTCC TCAAGGGAAA ACGCCTGTCT TTTGAAGCCA TTCAAACCTT	4140
CTCAACACCA CGTCGTTTGG CTGTTCTGTG AACTGGTCTT GCAGACAAAC AGTCTGATTT	4200
AACAGAAGAT TTCAAGGGTC CAGCAAAGAA AATTGCCTTA GATAGTGATG GAAACTTCAC	4260
CAAAGCAGCT CAAGGATTTG TCCGTGGGAA AGTTTGTACT GTTGAAGATA TCGAATCCG	4320
TGAAATCAAG GGTGAAGAAT ATGTCTATGT CACTAAGGAA GAAATTGGTC AAGCAGTTGA	4380
AGCCATTGTT CCAGGCATTG TGGATGTCTT GAAGTCACTG ACTTTCCTG TCAGCATGCA	4440
CTGGGCGGGA AATAGCTTTG AATACATCCG CCCTGTTTAC ACTTTAACTG TTCTCTTGGA	4500
TGAGCAAGAG TTGACTTGG ATTTCCCTGA TATCAAGGGA AGTCGTGTGA GTCGTGGCCA	4560
TCGTTTTTTG GGACAAGAAA CCAAGATTCA GTCAGCATTG AGCTATGAAG AAGACCTTCG	4620
TAAGCAGTTT GTAATCGCAG ATCCATGTGA ACGTGAGCAA ATGATTGTTG ACCAAATCAA	4680
GGAAATTGAG GCAAAACATG GTGTACGTAT CGAAATTGAT GCGGATTGTC TGAATGAAGT	4740
CTTGAATTG GTTGAATACC CAACTGCCTT CATGGGAAGT TTTGATGCTA AATACCTTGA	4800
AGTTCAGAA GAAGTCTTGG TGACTTCTAT GAAGGAACAC CAGCGTACT TTTGTGTTCTG	4860
TGATCAAGAT GGAAACTCT TGCCAAACTT CATTTCTGTT CGTAACGGAA ACGCAGAGCG	4920
TTTGAAAAAT GTCATCAAAG GAAATGAAAA AGTCTTGGTA GCCCGCTTGG AAGACGGAGA	4980
ATCTTCTGCG CGTGAAGACC AAAAATTGGT GATTTTCAGAT CTTGTTGAAA AATTAAACAA	5040
TGTCACCTTC CATGAGAAGA TTGGTTCTCT TCGTGAACAC ATGATTCGTA CGGGTCAAAT	5100
CACTGTACTT TTGGCAGAAA AAGCTAGTTT GTCAGTGGAT GAAACAGTTG ACCTTGCTCG	5160

507

TGCAGCAGCC ATTTACAAGT TTGACTTGTT GACAGGTATG GTTGGTGAAT TTGACGAACT	5220
CCAAGGAATT ATGGGTGAAA AATACACCCT TCTTGCTGGT GAAACTCCAG CGGTGGCAGC	5280
TGCTATTTCGT GAACACTACA TGCCTACATC AGCTGAAGGA GAACTTCCAG AGAGCAAGGT	5340
CGGCGCAGTT CTAGCCATTG CAGACAAAT GGATACGATT TTGAGTTTCT TCTCAGTAGG	5400
ATTGATTCCA TCAGGTTCTA ATGACCCTTA TGCCCTTCGT CGTGCAACTC AAGGTGTGGT	5460
TCGTATCTTG GATGCCCTTG GTTGGCACAT TGCTATGGAT GAGCTGATTG ATAGCCTTTA	5520
TGCATTGAAA TTGACAGTT TGACTTATGA AAATAAAGCA GAGGTTATGG ACTTTATCAA	5580
GGCTCGTGTT GATAAGATGA TGGGCTCTAC TCCAAAAGAT ATCAAGGAAG CAGTTCTTGC	5640
AGGTTCAAAC TTTGTTGTGG CAGATATGTT GGAAGCAGCA AGTGCTCTCG TAGAAGTAAG	5700
CAAGGAAGAA GATTTTAAAC CATCTGTGA ATCACTTTCT CGTGCCCTTA ACCTGGCCGA	5760
GAAGGCAGAA GGGGTTGCTA CGGTTGATTC AGCACTATTT GAGAATGACC AAGAAAAAGC	5820
TTTGGCAGAA GCAGTAGAAA CACTCATTTT ATCAGGACCT GCAAGTCAGC AATTGAAACA	5880
ACTTTTTCG CTTAGCCCAG TCATTGATGC TTTCTTTGAA AATACTATGG TAATGGCTGA	5940
AGATCAGGCT GTCCGTCAA ATCGTTTGGC AATCTTGTC CAACTAACCA AGAAAGCAGC	6000
TAAGTTTGCT TGTTTTAACC AAATTAACAC TAAATAAAT TTGATAAACG GACTTTATCT	6060
TATTACAAAG GAGAAGAAAT GGATCCGAAA AAAATTGCTC GTATCAATGA GCTTGCTAAA	6120
AAGAAAAAA CAGAAGGCTT AACACCAGAA GAAAAAGTGG AACAAAGCAA ACTACGTGAG	6180
GAGTACATCG AAGGTTATCG CCGCGCTGTT CGTCACCACA TTGAAGGAAT CAAAATTGTG	6240
GACGAAGAAG GAAACGATGT TACACCAGAA AACTACGCC AAGTACAACG TGAAAAAGGA	6300
TTACATGGCC GTAGTCTTGA TGATCCAAAT TCATAATAAT ACTCTTCGAA AATCAAATTC	6360
AAACCACGTC AGCTTCACCT TGCCGTACTT AAGTACAGCC TCGCGCTAGC TTCCTAGTTT	6420
GCTCTTTGAT TTTCATTGAG TATATGTATT CTTTCTTTTA ACAAAGATAG ATGAAACGAT	6480
AACAAAGAGA CTAGCAGTTT GTGTTTGCTA GTCTTTTTTC GCTAAAAAAG GAACCATAAT	6540
GGTTCCTAAA AACTATCATT AGTAACTTGC ACCGGCTGTA GCGTCTGCGT CACCACCGTG	6600
GCCTCCAGCA TCCCCTGAAT CAGAAGCGCC AGAAGTAGCA TCGGCGTCTC CATGACCTCC	6660
GGCAGCAGGA GCAAATGGTC CGCTACCACC CACCAAACGT TGACCAGTCT CTTTTAGGTA	6720
CCAGTCAAGC CATGGTTGGA AGTTAAAGAC GATTTTCATTG ATACCAGCGT ATGATCCATC	6780
AGGATAGTAC ATTGCTTGGT AGTTGTGAGT GTTGATAACA CCTGCAGGAG AACCTGGAAC	6840
GATCGTACGG ACGTATTCTT GGTTCGCTT GCGAAGTGTT CCGATAACCC ACTCTACGTT	6900

508

CTTCATACGT GCTGGTGGAA GAGAACCATG AACAGTCGAC ATACGGCTAC CTGATTGAGG	6960
TGGTACACGT TTAGCGAACA TAGTGTCTGG ATCTTGGTGA GCGTTGTTGT AGTAGAGGAA	7020
TTGGTTGTTG TCGTCAGCGT ATGTCAATTC AAATGGCATA GCTTTCAAGA ACATATCAAT	7080
TTGGTTAACT GTTAGGATAC CGTGGTCCAA TTTGACATAG GTATCACCAG AAACAGCACC	7140
AGTGAATGCT GCAACTTTTT CTACCCATTC TGGATCGTCA GGGTCAACTT CTGTGATGGT	7200
TGTAGCGATT GGTTTTCCAC AATCCAAGTC TTCTGATTCG ATTGGTTTGT GTTTTTTCAA	7260
TTTCGAAACG ACTCCTACGT ATTTAACAAA GTTATCTAAG CAAGTTTCAA GGAATTTAAC	7320
AGTGCCTTCG TTGGTGATAT TTCCGTTGTT ATCAAAAGCT TCCTTAGCTT TACCAAGAAG	7380
GAATTCGTTA CCTGGAAGCG TGTAGGCATT AACACCTGGA GCATCAAGGA TTTTACGAAG	7440
GTGAACCTGA GCACGTGATG TTCCTTGCTC ATAGTATGAT GCACCCACAA TCATAACAGG	7500
CTTGTTTTCA AATGGATGAA CTTCGTATGA AAGCCATTCA AGTACAGATT TGAGTGAAGC	7560
TGAGATAGTG TGGTTATGCT CAGGAGTAGC AATGATAACA CCATCTGCAC GAGTAATTTT	7620
GTTATATAAA TAACGTAATT GGAACTTTC ATCCCATTTT TCATCTTGGT TAAACATTGG	7680
AACTTCGTCA ATTTCAAGAA CTTCTAATTC AAATTTGAGT TTGAAGTAGC GACGGATAAA	7740
TTCCAAGAGC TTACGGTTAT ATGATTGATC GTAGTTTGAT CCAACAAGTC CAACAAATTT	7800
CATTCTTTTT GGTCTCCTAT CTTACAAATT TTCCAGTCA AAGTCTTCAG CATCTTGCG	7860
AAGTAATCT TGTGCATTAC GTAATTTTC TGTGATTTT ACAAGATAC GGAAGTCATC	7920
AAAGATGGCA TCCAATTTCT TGATAACATC AAGGTCAACC AAGTCGCCAC TTGGGTTAAA	7980
TGCTTGAAGA GAGTGTGAGA GCAAGAATTC ATCTGGAAGA ACATTTGCCT TGATTTCAGG	8040
AGCATTCAAG ATTTGACGAA GTTGCAATTG GGCACGAGAT GAACCAAGCG TACCGTAAGA	8100
AGCACCTGTA ATCATGATTG GTTTGTTCAA AAGTGGGTAA ATACCATAAG ACAACCAAGC	8160
AAGAGCGCTC ATCAAAACAG CTGGAATAGA GTGATCATAC TCAGGAGTAC CGATAATAAC	8220
GCCATCTGCC TCTTCGATTT TAGCAGCAAT TTCCAATATT TCAGCAGGTA CTTGCTTGTC	8280
AGCTGGTTTG TTGAAGACAG GAATGGCCTT GATTTCACA AGTTCAATTT CAGCTTTGTC	8340
AGTAAAGTGT TTTTGCATGT ATTGAAGCAA TTGACGGTTT GTAGAACGTT TTGAATTTGT	8400
TCCAACAATA GCAATAAGTT TTAACATGAG ATTTCCTTTC TCTTTTACA TAATACAATT	8460
TTAAAATTCC ATTGAAACAG TTGTCTCTAT AGAGTAGGAA TTCCTGAAGA ACAGCTTAGG	8520
TGGCCTTCTT TATCGATGAG GATGACTTCG ATGCCCTCCA AACTTTTCGAC TTGCCAGAGG	8580
ATAGAAGCAG GTCTTTCTCC AAAGAGTCGA GTCGTCCAGA TTTCCGCATC GACTGATTTA	8640
TCAGAGATGA TTGTTAGACT CGTAGTTCC GTTCAACAG GATATCCTGT TTGACTGTCA	8700

509

AAAATGTGAT GGTAATCTTG TCCATCGACG GTCAGGTGAC GTTCATAAAT GCCTGAAGTC	8760
ACGACAGATT TATTGACAAC AGGGATGGTC ATTAAATGAT TTCCCCTAGG ATTGGCTGGG	8820
TCTTGAATCC CGATTGCCA TGGGTATCC CCTCTTGCTT GATTTTTCCT AATGGTCAGG	8880
ATATTCCCTC CCAGATTGAT CAAGGCAGAA GTCACCCCTT CTTTCCTAAG AAATTGGGCA	8940
ACCTTATCCG CACTGTATCC TTTGGCTAAA CAACCTAGAT CGATCTTCAT TCCTTTCTGT	9000
TTTAAAAACA CAGTAGAAGT AGAAGAATCT AACTCGATAC CATGAGGATT GATTAGAGGC	9060
AGCACCGATT CAATTTCTTG AGGCTGGGCG ACCTTGGCAT CTGAAAACC GATACGCCAG	9120
GTTTGAATTA AGGGACCAAT GCTGATATTG AGGTGGCTAG AGAGCGCTAG GCTATGCTCT	9180
AACCCAAGTG AAATCAGCTC AAACAGGTCT GGATGAACCG TGACGGGGGC TATTCCTGCT	9240
TGATAATTGA TTTCCATCAA CTCAGATTCT TGACTATTGG CGTTGAAGCG GTATTCAAAGT	9300
TCTTTGAGCA AGTCAAAGGA TTTTGGGAGA AAGATATCGG CTTGCTCATC CACTAATGAA	9360
ATAGTGATAG TAGTCCCAT TAGCCGTTC GAATGTGAAC GAAGAGTCAA GCTACCAACT	9420
CCTTTCTCTT ATAGAAAATA AGTTGTAATA TCAAATAATC ATCTAAATTG AAGCCCTTAC	9480
ATTTTCATTTT CATGTTATTA TAATACCATA AAGTTAGAAT TTTCAACAAC AAAATTGGA	9540
AAAAGTCAAG AAATATGCTC ATAAAATTCA TCAGGCTTGA AAACAGGATA AATGGGGAAT	9600
TATTTTGTAT AAAAAATGCT GAAATAATAG TACCCCTCTT GTAAACGCTA ACGGTAAATG	9660
GTATACTAGT AAGGTAAATT TAGAATGAAG GCAGGAAATT TTTATGAGTA AAATCGTTGT	9720
AGTCGGTGCT AACCACGCTG GTACAGCATG TATCAATACC ATGTTGGATA ATTTTGAAA	9780
TGAGAACGAA ATTGTTGTAT TTGACCAAAA CTCTAACATC TCTTTCCTAG GATGTGGAAT	9840
GGCTCTTGG ATTGGTGAAC AAATTGACGG TGCTGAAGGC TTGTTCTATT CTGATAAAGA	9900
AAAATTGGAA GCTAAAGGTG CTAAAGTTTA CATGAACTCA CCTGTTCTTT CAATCGACTA	9960
TGATAACAAA GTAGTTACAG CGGAAGTTGA AGGAAAAGAG CACAAAGAAT CATAAGAAPA	10020
ATTGATTTTC GCTACAGGCT CTACACCAAT CTGCCCACCA ATCGAAGGTG TGAAATTTGT	10080
TAAAGGAAAC CGCGAATTTA AAGCAACTCT TGAAAACGTA CAATTCGTGA AATTGTACCA	10140
AAATGCTGAA GAAGTTATCA ATAACTTTTC TGACAAGAGC CAACACCTCG ACCGTATCGC	10200
CGTTGTTGGT GGTGGTTACA TCGGTGTTGA ACTTGCTGAA GCCTTTGAAC GTCTTGAAA	10260
AGAAGTTGTC CTGTTGATA TCGTTGATAC TGTCTTGAAC GGTACTATG ACAAAGACTT	10320
CACACAAATG ATGGCGAAGA ACTTGGAAGA TCACAACATC CGCTTGCTC TAGGTCAAAC	10380
TGTTAAAGCA ATCGAAGGTG ACGGTAAAGT TGAACGCTTG ATTAAGTACA AAGAAAGCTT	10440

510

TGACCTGGAT ATGGTTATCC TTGCAGTTGG TTTCCGTCCA AACACAGCCC TTGCAGGTGG	10500
TAAGATCGAA CTCTTCCGCA ACGGTGCCTT CCTTG TAGAC AAGAAACAAG AAACATCTAT	10560
CCCAGACGTT TACGCTGTTG GTGACTGTGC GACTGTTTAT GACAATGCTC GTAAAGATAC	10620
AAGCTATATC GCTCTTGCTT CAAATGCTGT GCGCACTGGT AACGTTGGT	10669

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7542 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CGCGCTAATA GATACTTTAT GATAGAATAA AGAACAAGAT TGACAAGTAA GAGGAAACAT	60
TATGCAAAAT CAAACACTCA TGCAATACTT TGAATGGTAT CTGCCCCACG ACGGTCAACA	120
CTGGACGCGT CTGGCTGAAA ATGCTCCACA CCTAGCTCAT CTGGGGATCA GTCACGTCTG	180
GATGCCACCA GCCTTCAAGG CAACCAACGA AAAAGATGTC GGCTATGGGG TCTATGACTT	240
ATTTGACTTA GGAGAGTTCA ACCAAAAAGG GACTGTCCGC ACCAAGTATG GTTTCAAAGA	300
AGACTATCTT CAAGCCATTC AAGCCCTTAA AGCACAGGGA ATTCAACCTA TGGCCGATGT	360
AGTTCTCAAC CACAAGGCTG CTGCCGATCA CAGGGAAGCC TTTCAGGTTA TCGAAGTTGA	420
TCCTGTAGAC CGTACAGTTG AACTTGGAGA ACCCTTCACC ATCAATGGCT GGACTAGTTT	480
TACCTTCGAT GGTGCGCAAG ATACCTATAA TGGCTTCCAC TGGCATTGGT ACCACTTCAC	540
CGGTACAGAC TACGATGCCA AACGCAGTAA ATCTGGGATT TATCTGATCC AAGGGGACAA	600
CAAGGGCTGG GCCAACGAGG AATTGGTCTGA TAACGAAAAC GGAAACTACG ACTACCTCAT	660
GTATGCCGAC CTAGACTTTA AACATCCTGA AGTCATCCAA AACATCTATG ACTGGGCTGA	720
TTGGTTCATG GAAACGACTG GTGTAGCTGG TTTCCGTTTG GATGCCGTTA AGCATATTGA	780
CTCTTTCTTT ATGCGCAACT TCATCCGCGA TATGAAGGAA AAATACGGTG ACGATTTCTA	840
TGTTTTTGCT GAATTTTGGA ACCCAGACAA GGAAGCCAAT CTGGACTATC TCGAAAAAAC	900
GGAAGAACAC TTTGACCTTG TCGATGTTCT TCTCCACCAG AATCTCTTTG AAGCCAGTCA	960
AGCTGGCGCA AACTATGACC TTCGTGGCAT TTTCACAGAT AGCCTGGTTG AACTCAAGCC	1020
TGACAAGGCT GTGACTTTTG TCGACAACCA CGATACCCAA CGAGGACAAG CCCTTGAGTC	1080
TACCGTTGAA GAATGGTTCA AGCCAGCAGC CTATGCCCTC ATTTTGTAC GCCAAGACGG	1140
CCTTCCATGT GTCTTTTACG GAGACTACTA TGGGATTTCA GGGCAGTATG CTCAAGAAGA	1200

TTTCAAAGAA ATCCTTGACC GCCTCCTAGC CATCCGAAAA GATTTCGCCT ATGGAGAACA	1260
AAATGACTAC TTGACCATG CTAACGTAT CGGTGGGTA CGTTCAGGTG CTGAAAATCA	1320
ATCCCCAATC GCAGTCCTTA TCTCAAATGA CCAAGAAAAC AGCAAGTCAA TGTTCGTCGG	1380
TCAAGAATGG ACTAATCAAA CCTTTGTAGA TTTACTTGGT AACCACCAAG GTCAAGTTAC	1440
AATTGATGAG GAAGGTTATG GACAATCCC TGTCTCAGCT AGATCCGTAA GTGTCTGGGC	1500
AGTCAATACC ATCTAATAGC TCATAATAAC CAAGCTAGGT CCAAGCGGAT TTGGCTTTT	1560
TGTATTCACA AAAAGACCTA CCCAAATGGA TAGATCTTTA CTTGATTACA ATTTACCTGC	1620
TACTGCATCC AACAATCTT GGATCTTAGG TTGGTTGCTT CCTCCTGCCA TGGCCATATC	1680
TGGTTTACCA CCACCACGTC CATCGATGAT TGGTGCTAAT TCTTTGACAA GGTTCCTGC	1740
ATGAAGGTCT TTTGTCTTGC TTGCTACAAG GACATTGACT TTGTCACCGA TAGCGGCAAC	1800
TAGGACAAGA AGATCAGAGT AGTCTTTTGG TTTCCAGTTA TCTGCAAAAG TACGAAGGGC	1860
ACCGGCATCG GATACAGACA CTTGACTAGC AATGTAACGA TGACCGTTGA CTTCTTAAC	1920
ATCTTTGAAG ATATCGCCTG CGGCTGCAGC TCGGGCTTTT TCTTTCAACT CAGCATTTTC	1980
TTTTTTGAAGT TGACGAAGTT GTTCTTGAAG TCCTTCTACC TTGTGAGGTA CTTCTCTGAC	2040
TTGAGGTGCT TTCAAGGTTG CTGCGATAGC TTAAAGAGCA TCCTCTTGT CACGATAGGC	2100
TTCAAAGGCT TCCTTACCAG TCACTGCCAA GATACGGCGA GTTCCTGAAC CGATTCCTTC	2160
TTCTTTGACA ATTTTGAAGA GACCAATCTC AGAAGTGTG TCAACATGAG TACCACCACA	2220
AAGTTCAATA GAGTAGTCAC CGATAGTCAC GACACGAACT TCCTTGCCGT ATTTCTCACC	2280
AAAGAGGGCC ATAGCTCCCA TTTCTTTAGC AGTGTCAATA TCCGTTTCAA CTGTCTTCAC	2340
TTCAAGTGCT TCCCAAATTT TCTCGTTAAC TTGCTGTTCA ATCGCACGAA GTTCCTCAGC	2400
AGTTACTGCT TGAAGTGGG TAAAGTCAAA GCGAAGGAAT TCAACTTCGT TAAGAGATCC	2460
TGCCTGTGTT GCGTGGTTTC CAAGGATATT GTGAAGGGCA GCGTGAAGCA AATGAGTCGC	2520
AGTGTGGTTT TTCATGACAC GGTGACGGCG ATTGCTATCA ATTGCCAAGG TATATCTTTG	2580
GTTCAGGCA AGCGGTGCAA GGACTTCAAC TGTATGAAG GCTTGACCAT TTGGGGCTTT	2640
CTGAACATTG GTCACAGTAG CCACAACCTT ACCTGACTCA TCCAAGATTT GTCCGTAGTC	2700
AGCTACCTGT CCACCCATTT CAGCATAAAA TGACGTTTCC GCAAAGATAA GAGAGGCAGT	2760
TCCTTCTGAA ACAGCTCCTA CTTCTGCATT GTCAGCAACG ATAGCTACCA ATTTAGAAGA	2820
CAATTGGCTA GCATTGTAGT TGAAGACACT TTCTACAGTG ATGTTTTGAA GAGTTTCATT	2880
TTGCATACCC ATTGAGCCAC CCTTGACAGC TGACGCACGC GCGCGTTCTT GCTGTTCTTT	2940

512

CATGGCTGCT TCAAAACCTT CACGGTCTAC AGTCATACCA GCTTCTTCAG CGATTTCCTC	3000
AGTCAATTCA ACTGGGAACC CATAAGTATC ATAGAGTTTG AAGACATCTG AACCAGCGAT	3060
AACAGATTGA CCTTTTCTTT TCAAGTCTGC TACAATGCCT TGGGCAAAGT GTTGACCTGA	3120
GTGAAGGGTA CGGGCAAATG APTCTTCTTC GCTCTTAACG ATTTTCTCAA TAAAGTCACG	3180
TTTCTCAAGC ACTTCTGGGT AGTAGCTTTC CATGATTTTT CCAACAGTTG GAACCAATTT	3240
GTAAAGGAAA GGCTCGTTGA TACCCAATTT TTGACCATGC ATAGAAGCAC GACGGAGAAG	3300
ACGACGAAGA ACATAACCAC GACCTTCATT TCCTGGAAGG GCACCATCAC CGATAGCAAA	3360
TGAAAGAGAA CGAATGTGGT CTGCGATAAC CTTGAAGCTC ATGTTGTCGC CATCTTGGTC	3420
ATAAACCTTA CCAGACAATT TCTCGACTTC ACGGATAATC GGCATGAAGA GTCCGCTTTC	3480
AAAGTTGGTC TTAGCCCCCT GGATAACGGC CACCAAACGC TCCAAACCAG CGCCCCGTATC	3540
AATGTTCTTA TGTGGCAATT CCTTGATTC GCTACGAGGA ACAGCAGGGT CTGCGTTAAA	3600
TTGTGACAAA ACGATGTTCC AGATTTCAT ATAACGGTCG TTTTCAATAT CTTCTGCAAG	3660
CAGGCGAAGA CCGATATTTT CTGGGTCAA GGCTTCCCCA CGGTCAAAGA AGATTTCTGT	3720
ATCTGGTCCA GAAGGTCCCG CACCGATTTT CCAGAAGTTG TCCTCAATTG GAATCAAGTG	3780
ACTTGGATCC ACTCCCACTT CAATCCAGCG GTTGTAAGAA TCTTTATCGT CTGGATAGTA	3840
GGTCATGTAA AGTTTTTCAG CAGGGAAATC AAACCATTCA GGGCTTGTC AAAGCTCATA	3900
AGCCCAAGTG ATAGCTTCGT CACGGAAGTA ATCCCCGATA GAGAAGTTCC CCAGCATTTT	3960
AAACATGGTA TGGTGACGCG CGGTCTTCCC TACGTTTTCG ATGTCGTTGG TACGGATAGC	4020
CTTTTGGGCA TTGGTAATAC GTGGATTTTC AGGGATAATG GTCCCCGTAA AGTATTTCTT	4080
AAGGGTTGCT ACCCCAGAGT TGATCCACAA AAGAGTTGGG TCATTTACAG GAACCAAAT	4140
TACTGATGGT TCTACTGAGT GACCTTTGGT CGCCCAGAAA TCAAGCCACA TTTGGCGTAC	4200
TTGTGCACTA GATAGTTGTT TCATATTGTC TCCTTATTC CTTGTTTAAT GTGATTGGCT	4260
TTCCAGCATT TCCACATAGT CAATCGCGAC ACAGAGGGAA ATGACTAGGT CTGCATAAGC	4320
GTCTTCAAGA ACCGTTACGG TATAGGTAGA AGTCAGATGG AAGAGTTCCT TCTTAATTTT	4380
CGCAATCAAC TGATCGCGAT CATCCAGCAA TTTGAAATTC AAATCCCAGA TATTGCCCTC	4440
GATACGAAGA CCTAGATTAT CAAACTCATA CTTATCTCGC CAGAAGGTCA ACTTCTTACG	4500
AATGACAAAA CTCGAGCCAT CCCGAAGCTG AATTTCAAAA CGAGGAAGCA AGGTCAAGAT	4560
TTCTTTACTA ATCTCACTGA CTTGTTTACC AGCCGCATCA TAGATGGTAA AGGTTTTAGG	4620
AATCTTAAAA AATGATCCCT CCACCTGATA GGCAATTTCT CCCCTGTCAT CCTTGATAGC	4680
GAAGCGTTCG CCTCCAAGAC GAACTTTTG TTTGACAAGA AATGTTTICA TCAACACCTC	4740

513

CAAAAATCAA AAGACAAGCT CATATCACGA AGGGCGAAAA ACCGCGGTAC CACCTTCATT	4800
CAATGAACTT GTCATTCTCT TGTTCTTATG CAATTGTATG ATTGAGTAGC ATGACTTCCT	4860
AGCTTAGATG GCTCGCAGCA CCGCCATTTC TCTGGACTAA GACAAGTGAA AATCAATTCT	4920
CAACTTTCTT ATTATAACGT TTTTPIAAGC TTGCGTCAAC TGGAAATGAT CTCCGTTGAA	4980
TTAGACCAAT TCCCTACATC TCTGATTACT TTTTCAGGAT ATATTTTTTC TTAGTGCCAT	5040
TTTTCTTTTT ATCCCAAATT TTCATATTAC TAAACACAGC TACTAGAATA TTTCCAAATA	5100
TAAAGGTGCC TATCACCCAA TATATGGACT CAGTTGTTAG GTATTGTGCGA TCCAAGCCAT	5160
CCTTTAAATG GAATAGTATA GCAGTTTGGT TAACAATCAT AAAGGTTGGC CAGAACTTT	5220
TTTGTAAAAA AGTAGACATT TTCATTATTT GTTGCCGCTT TCTGTAAGGT TAATACTCAA	5280
TAAAAATCAA AAAGCAAACCT AGGAAGCTAG CCTCAAGCTG TACTTGAGTA CGGCAAGGCA	5340
ACGCTGACGT GGTTTGAAGA GTATAGGCTT AGTATACTAC TAGGCAAGCA AATAAACAAA	5400
TAAACAATA GAATAGAAAA AGATAGGGCT CTAAAACTG ACTTCTATTC CTTAAAAACG	5460
AACCAGCTTG ACTGATTCGT CTCTTACGT TTATCTCCTA CTTCCGATAC ATTTTAAACT	5520
GTAGGAAGAG GTCGCTATAT TTCCCTGTCC ATTTATGGTC AAATTTCTCA TAAACTTCTA	5580
GGTGTTCAT GGTTCACAAC TCGGGATAGA AGGCCTTATC TTCCCTTGTG TCCTCTGGGA	5640
GCAATTCCTT CGCTGGTAGG TTTGGTGTG AATAGCCGAC ATACTCCGCA TTTTGGAGAG	5700
CATTTTCAGG TTTCAACATA AAGTTGATAA AGGCATAGGC TGAGTTTGG TTTTAACTG	5760
TTTTGGGAAT GACCATATTG TCAAACCAA GATTGCTGGC CTCTGTCGGT ACCACATAAC	5820
GTAGATTTTC ATTTTCTTCT AACATTTGGC TGGCTTCACC AGAGAAGGTC ACGCCGATTG	5880
CAACATTATT CTGAATCATA TAGCCCTTCA TCTCGTCCGC AACGATAGCC TTGATATTG	5940
GAGTCAGTTT GTAGAGCTTA TCCACTGTCT CTTCCAACTG CTGCAGATCC TTGGAGTTGA	6000
GGCTGTAGCC GAGGGAATTG AGTCCTAGTC CCAGCACCTC ACGCGCCCCA TCAAAGAGCA	6060
TGATAGAATT CTTATACTCC GGCTTCCAAA GGTCAATCCA ATGCTCAGGC GCTTCATCTA	6120
CCATGGTTTC GTTGTAGACA ATTCCTAAGG TTCCCAGAA GTAAGGGATG GAGAATTTAT	6180
TACCTGGGTC AAAGGACTGG TTGAGAACT CTGGTCCGAT ATTTTCGATT CCTTCAATTT	6240
TTGAATAATC AAGCGGAACC AAGAGGTCTT CGTCCTTCAT CTTGTTAATC ATGTATTAC	6300
TTGGAATGGC AATATCGTAG GTCGTTCCAC CCTGCTTTAT CTTAGTGAC ATGGCTTCGT	6360
TGGAGTCAA AGTCTCGTAC TGAACCTGAA TTCTGTTC TTCTGTAAAC TGAGTCAAGA	6420
GTTCAGGATC GATATAGTCT CCCAGTTAT AGATAACCA TTTTGTACTA TCTCGACTAT	6480

514

TGATTTTACT ATCTAAATGA GTCGCAATTC CCCACAAGAC AAGGATAATC GCTGCAATTC	6540
CTGCTAAAAA TGAATAGATT TTTTTCATGC TTGCTCCTCC TTCTCACGAG AGATAAAGTA	6600
ATAACCTACA ACTAGGATAA TACTAAAGAG AAAGACTAGA GCAGACAGGG CATTGATTTC	6660
TAAGGAAATC CCCTTGCGAG CACGAGAGTA AATCTCGACT GATAGGGTTG AAAAGCCATT	6720
TCCTGTTACA AAGAAGGTCA CGGCAAAGTC ATCTAACGAA TAGGTGAAGG CCATGAAATA	6780
ACCAGTAATG ATAGACGGAG TCAGGTAAGG AAGCATGATT TCCTTGAACA TCTGAAATTG	6840
ACTAGCTCCC AAGTCATAGG CCGCATGAAT CATGTCGCCA TTCATTTCCT TGAGTCGAGG	6900
CAAGACCATC AAGACCACGA TAGGAATGGA GAAGGCCACG TGACTAGATA GAACGGTCAA	6960
AAAGCCAAGT GAAAACCTGA GTTGGGTAAA GAGAATCAAG AAGCTAGCAC CAATCATAAC	7020
GTCAGGCGCA ACCATGAGGA TATTATTGAG TGATAGAAAAG GCTTCTTGGT ATTTCTTACG	7080
AGACTGGTAG ATGTAAATGG CACCAAAAGT CCCGATAATG GTCGCTATCA AGGCTGATAG	7140
GAAGGCCAAG AAAAATGTCT GAGCCAAAAT CAGCATGAGT CTCCCATCTC CAAACATGGT	7200
TTCAAAGTGA GTCCAGCTAA AACCTGTAAA GCTATTTCATA TCATCACCAG CATTAAGGC	7260
ATAGCCAATC AAGTAAAAGA TAGGCAGGTA GAGGACCAGA AAGACCAGTC CCAGATAAAG	7320
GTTGGCAAAT TTTTTCATCG TTCTCTCCTT TCCTTAGTCA CCCACATGGT GATGAACATG	7380
GTCAGGATGA GAATCACACC GATGGTTGAA CCCATACCAT AGTTGTCATT GGTTAGAAAA	7440
TTCTGTCTAA TAGCCGTCCC CAAGGTGATA ACGCGTTCCC ACCAATCAAA CGGGTCAGCA	7500
TGAAGAGACT CAAACTTGGG ATAAAGACCG ACTGAACCCC GG	7542

(2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9223 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAACCAAAT TCCGGTATTT TAACCTATGC TGTAATACC ATGAAGTCTG TCATGACAGA	60
TCAGGTCTAT AACATTAAGG TTGAGACAGA AAATGGAAAT TATGTTGGTG AAGCTAGCCA	120
TGTTTGGTC CTTTGTACAA ATTACTTCGC TGATAAGAAA ATCTTTGAAG AAAACAAGGA	180
CGGCTATGCC AACATTTTGA TTCTGAAAGA TGCCTCTATA TTCTCCAAAT TATCCGTCAT	240
TCCTGATTTA TTAAGGGG ATGTTGTCGC AAATGATAAT ATCGAGTATA TCAAAGCGCG	300
TAATATTAAA ATCTCTCAG ATAGTGAATT GGAGTCAGAT GTTGACGGAG ATAAATCAGA	360

515

TAACCTACCT GTAGAAATCA AAGTCCTAGC TCAGCGAGTA GAAGTATTTT CAAAACCGAA	420
AGAGGATTAG TATATAGAGA AAGCCTTTTT TAAGGCTTTT TGTATACTTT AAAAGATAGT	480
TCCTTTAACA ACGGACATTC CTTGCAAATA GTTTTACAAA AATAGTATAC TGGATTCAAT	540
GAGTTTGAAA ACGTTTTCGT AAAATTTGAA TGAATACTTT AGGAGACAAA TTGATGGAAT	600
TGAGTGCTAT TTACCATAGG CCTGAGTCGG AGTATGACTA TCTTTATAAG GATAAGAAAC	660
TCCATATTCG AATTCGAACT AAGAAAGGGG ACATTGAAAG CATCAACTTG CACTATGGGG	720
ACCCTTTTAT CTTTATGGAG GAGTTTTATC AGGATACAAA AGAAATGGTC AAGATAACTT	780
CTGGTACCTT ATTTGACCAT TGGCAGGTG AAGTGTCAGT TGACTTTGCA CGTATCCAGT	840
ATCTCTTTGA GCTCAGAGAT ACAGAAGGTC AAAATATTTT GTATGGCGAT AAAGGGTGTG	900
TGGAAAATTC TCTAGAAAAT CTTTCATGCA TTGGGAATGG ATTTAAGTTG CCTTAGCTTC	960
ATGAGATTGA TGCCTGCAAG GTTCCTGACT GGGTTTCAA TACGGTATGG TATCAGATAT	1020
TTCTGAAAG ATTTGCCAAT GGCAATGCTC TATTAAACCC AGAAGGGACT TTAGACTGGG	1080
ATTCATCTGT CACACCTAAG AGCGATGATT TCTTTGGTGG TGATTTACAG GGGATTATTG	1140
ATCATATGAA TTAATTGCAA GACTTGGGTA TTAAGGACT ATATCTTTGT CCCATCTTTG	1200
AATCTACAAG CAATCACAAG TACAATACGA CAGATTACTT TGAAATTGAC CGTCATTTTG	1260
GAGACAAGGA GACCTTTCGG GAACTGGTGG ATCAAGCGCA TCATCGTGGC ATGAAAGTCA	1320
TGCTGGATGC GGTATTTAAT CATATTGGTT CGCAATCTCT TCAATGGAAA AATGTCGTCA	1380
AAAAAGGTGA ACAGTCTGCT TATAAGGATT GGTTCATAT TCAACAATTC CCAGTGACAA	1440
CTGAAAAGCT AGTTAATAAG AGAGACTTAC CCTATCATGT TTTTGGTTTC GAGGACTATA	1500
TGCCTAAGCT AAATACAGCC AATCCAGAGG TCAAGAATTA TCTTTTAAAG GTTGGGACTT	1560
ATTGGATTGA AGAGTTAAT ATCGATGCTT GGCCTTTGGA TGTGGCTAAT GAGATTGACC	1620
ATCAGTTCTG GAAGGATTTT CGTAAGGCAG TTTTAGCTAA AAATCCTGAT CTTTATATCC	1680
TAGGAGAAAT CTGGCATACA TCTCAGCCTT GGCTAAATGG AGATGAGTTC CATGCCGTCA	1740
TGAATTATCC TTTATCTGAT AGTATCAAGG ACTATTTCTT ACGAGGAATT AAGAAGACAG	1800
ACCAGTTCAT CGATGAAATC AATGGAGAGT CTATGTATTA CAAGCAGCAG ATTTAGAGG	1860
TCATGTTTAA TCTCTTGGAT TCACATGATA CAGAGCGAAT CCTGTGGACG GCCAATGAAG	1920
ATGTTCAACT GGTAAATCA GCCTTAGCCT TTCTCTTTTT AAAAAAGGA ACACCGTGCA	1980
TTTATTACCG AACCGAGCTA GCCTTGACTG GAGGACCAGA TCCAGATTGT CGTCGTGTA	2040
TGCCTTGGGA ACGTGATCA AGTGACAATG ATATGCTGAA CTTTATGAAG AGGCTGATTA	2100

516

AAATTCGGAA ATACGCGTCA GTAATCATTT CGCATGGCAA GTATAGCCTT CAAGAAATCA	2160
ACTCTGATCT AGTAGCTCTG GAATGGAAAT ACGAAGGACG GATCCTCAAA GCAATATTCA	2220
ACCAATCAAC AGAAGATTAT CTTTATAGAGA AAGAAGCAGT AGCACTAGCA AGCAATTGCC	2280
AAGAATTGGA TAATCAGCTT GTCATCTCTC CAGATGGATT TATGATTTTC TAAAACTAG	2340
TTGATGAAGA TTATGGTACA TTTCATACCT TATATAGTAT AATAAGGCTA GTTACTAAAC	2400
TTGTAAAGGA GAACTTAAAT GAATTGTAGA GGACATGAAA CAAGACAAAG AATTGTTAGA	2460
GATTTTGAAG TTCAGCCTAA AGCACATATT AAGCTGTTAG CAAATCAACA AAAACATAGT	2520
GATGCAGGAG CAACTATTGA AGATGAATAT TATGTATTTA TCGCTGAGAG TAAAAATGAT	2580
GGCAAGAAGG AAGTTATTCA GTGTTGCATG GGTGCGGCAA GGGATTTTTT AGAACTAATT	2640
AATCACAAAG GGCTACCTCT TTTTAATCCG CTTGTAGGTG ATTCTCATGT AAATAATAGA	2700
CAAGAATATG ACAATACAGG GAGTGGAAAT TTATAACCTG AAAAGTGGAA TGAAACTGCA	2760
AAGCAGCTTT ATAATGCTAT AATGTGGTTG ATTATTTTAT GGAATGCTAA GCCGGATACA	2820
CCTTTATTTA ATTTTAAAGA CGAAGTAATT AAGTATAAAA CATATGAGCC TTTTGAAAGC	2880
AGTATAAAA GAGTAAATAC TACTATAAAG AATGGTAGTA AAGGAAAAC TCTGACTGAG	2940
ATGATTAATG GCTACAGAGC GGATAACGAT ATTAGAGATG AAATTGTAA CTTTAATATT	3000
CTGAAAAATA AAATTCGTGA TATGAAAAAC CAACAAGGAA ATACAATGGA ATCTTACTTT	3060
TAGTTATTGT TGAATTTTGG GTATTCTATA AAATATCCTA ATTGAGATTT AAATAGTAGA	3120
CTATACAATA TAGTTAAAT ATCAGTAAAA ACAACACTTT ATTGAGGTAT TGGATACGCT	3180
TTGCTAATAG CCTAATAATC ACATGTGGAG TGTGCTACA ACGAAAAAGG TGATAATCCT	3240
TGATTTCAAG CTATTTTATA AGCATTTTGT CTTGTAGAT AAAGGCAATT TTGACAATAA	3300
AAATCCTAAA AGGTGAATCG TTATAGATGT ATTTGTAGAT ATCGTTTGCG CATCGAAAAA	3360
ATTAATACAA GAATAAATAT TTATAGCTCT TTAGGTGACT TTTATAGAAG TAAAGTTTAG	3420
GATAGAAAA CAAGAAATAA CGCACCATTT TTGGTGCGTT ATGCTTTTTT ATGCTATAAT	3480
GGATTATATA AAATAAAGGA GTTTGCTATG ATTGGAAGA ACATAAAATC CTTGCGTAAA	3540
ACACATGACT TAACACAACG CGAATTTGCA CGGATTGTAG GTATTCACG AAATAGTCTG	3600
AGTCGTTATG AAAATGGAAC GAGTTCAGTC TCTACCGAAT TAATAGACAT CATTGTCAG	3660
AAGTTTAATG TATCTTATGT CGATATTGTA GGAGAAGATA AAATGCTCAA TCCTGTTGAA	3720
GATTATGAAT TGACTTTAAA AATTGAAATT GTGAAAGAAA GAGGTGCTAA TCTATTATCT	3780
CGACTCTATC GTTATCAAGA TAGTCAGGGA ATTAGCATTG ATGATGAGTC TAATCCTTGG	3840
ATTTTAATGA GTGATGATCT ATCTGATTTG ATTCATACGA ATATCTATCT AGTAGAAACT	3900

517

TTTGATGAAA TAGAGAGATA TAGTGGCTAT TTGGATGGAA TTGAACGTAT GTTAGAGATA	3960
TCTGAAAAAC GGATGGTGGC CTAATGGAAA TCCAAGATTA TACTGATAGT GAATTCAAAC	4020
ATGCTTTAGC AAGGAATCTT CGTTCCTGA CAAGAGGAAA AAAGTCCAGT AAGCAACCTA	4080
TAGCGATTTT GCTTGGAGGG CAAAGTGGTG CCGGTAAGAC TACAATTCAT CGTATTAAAC	4140
AGAAAGAATT TCAAGGAAAT ATTGTTATCA TAGATGGTGA TAGTTTTCGT TCTCAGCATC	4200
CACACTATTT AGAACTGCAG CAAGAATATG GCAAAGACAG TGTAGAATAT ACCAAAGATT	4260
TTGCAGGAAA AATGGTAGAG TCTTTAGTAA CAAAATTGAG TAGTTTGAGA TACAATCTTT	4320
TGATAGAGGG AACTTTACGA ACAGTTGATG TTCCAAAGAA AACAGCACAA CTCTTGAAAA	4380
ATAAGGGATA TGAAGTACAA TTGGCCTTAA TTGCGACAAA GCCTGAATTG TCGTATCTAA	4440
GTACTCTTAT CCGTTATGAA GAACTGTACA TTATCAATCC AAATCAAGCA CGCGCAACTC	4500
CAAAGAACA TCATGATTTT ATTGTAAATC ATCTAGTTGA TAACACACGA AAATTGGAAG	4560
AACTAGCTAT CTTTGAAAGA ATTCAAATTT ACCAACGAGA TAGAAGTTGT GTATATGATT	4620
CAAAGAAAA TACAACCTCA GCAGCAGATG TTCTTCAAGA GTTACTCTTT GGGGAGTGGA	4680
GTCAGGTAGA GAAGGAGATG TTGCAGGTGG GGGAAAAGAG ACTTAATGAA TTACTTGAAA	4740
AATAACAAT TGATATTTTT AGGAGAATAG AAATGAGAGG GTTTAATAAC AAGATAAAGT	4800
CTGTTTATCA AGAACTAACA AATTCCAAAG AGAAATTCGG TAGCTTTTAC AAGACTTTAA	4860
TTCAATTTGCA TACACCTGTT TCTTATGATT ACAAGCTATT TTCTAATTGG ACTGCAACGA	4920
AATATAGAAA AATTACTGAA GATGAACTAT ATGATATATT TTTTGAAAAT AAGAAAAATA	4980
AAGTTGATAA GACAATTTTT TTTAGTAATT TTGATAAGGT TGTTTTTTCT AGTTCAAAAG	5040
AATATATTAG TTTTCTTATG TTAGCAGAGG CAATCATAAA AAATGGAATA GAAATAGTTG	5100
TAGTAACTGA TCATAATACT ACCAAAGGTA TTAAAAAGTT ACAAATGGCA GTCTCAATCA	5160
TAATGAAAAA TTATCCGATT TATGATATAC ATCCTCATAT TTTACATGGA GTAGAAATTA	5220
GTGCAGCAGA TAAATTGCAT ATTGTATGTA TATATGATTA TGAACAAGAA TCATGGGTTA	5280
ATCAATGGTT AAGTGAAGAT ATTATAAGTG AGAAAGATGG AAGTTATCAA CATTCCTGA	5340
CTATAATGAA GGATTTCAT AATCAAAAA TAGTTAACTA TATTGCTCAT TTCAATAGTT	5400
ATGACATTTT GAAAAAAGGT TCTCACTTAT CAGGTGCATA TAAACGAAAA ATTTTTCCTA	5460
AAGAAAAATC ACGATTTTGG AGTTTAATAT TAACTCGAAA GAATCTTCGC AACAACTGA	5520
TATTCTCTAT AAAGAAGTTG GTGTATTAAG TTTGGGACAA AAAGTTGTAG CCATGCTTGA	5580
TTTTTTATTA GCATATAGTG ATTATTCTAA AGACTTCAGA CCATTGATTA TTGATCAGCC	5640

518

TGAAGACAAT CTAGACAATC GTTATATTTA CAGGCATTTA GTTCAGCAGT TTAGAGATGT	5700
GAAAGCTCAA CGTCAAATTA TTTTAGCAAC ACATAATGCT ACAATTGTAA CAAATTCCTAT	5760
GACAGATCAA GTTGTTATTA TGGAGTCAGA TGGAGTTAAC GGATGGATTG AATCACAGGG	5820
ATATGTTAGT GAAAAATATA TAAAAAATCA TATCATCAAT CAATTAGAGG GAGGAAAAGA	5880
TTCTTCAAG CATAAAATGT CTATATATGA GACGGCTTTA TCAGAGTAGA GTCAGAAAAA	5940
GTAGGTTAGA AATTTAGCCT ACTTTTTTCT TTGTCCGACA GGCATAGTGT ACATCTGAGG	6000
TCCAAGTCCT CTGTGGATAT TTGCTGCAGA TGAAACCAAT AGCGACTCCT AAGCCTGAAT	6060
ATCGTGAGGT AGGGGGGATA GGAAGGAATT AGCGAAATCA AGGTTCTACA AACAGAATCG	6120
TGACTTGAAG CCATATATAG CGGATGAGGA ACTCTAAAAT CCAAATAGGT GTCGTAACCT	6180
ATATACGTAA ATTACGAGAG TAAACTAGGA AAGATGTACG GCTTATTCCTG TGAGCGTTTA	6240
GGACGTAGTA CAACGAATCA TGGGAGTCAG CTGAACACAT AGTATTGAAG AAATTTCTGT	6300
AATGGAAATG GAGCGAAGAA GTGAACAATT AAATGAATAC CTCTCTAATT AAATTTGTCA	6360
ATTCTAATTC CTGGTATGAA AAGACAGTGA CCTGAAAATG TAAACGATGG GAGCTGATCA	6420
TAAATATAGG ACGGTACATG CAGTGGTGT AGAGATTAGT CCTTACTTGA TTTGTGATAA	6480
CTTCCCCAAA TTTCTCTGCG TATACTTTTC TCAACTTTTA AAAATCCAAC TAAGAATTTT	6540
ACCTGGGGGT TTGGGGGCGG AGCACTAAGT TATCTTATCG TTAGCTGTCA AAAGTGGTAG	6600
GTTTTGATAG GCTGGCGATA TGATTTTGG GATATTGTGG ACACAATATC TGAGCTCGCA	6660
AAGCCTTACA AGAATGAAAA TCAGTTGTTG GAAAAGTGT CTGACATTGT ATGGTAGCTC	6720
ACATTGTCAG TACAAGTATT TTGGAAGGA AGTAGCAGTA TGAAACGAGA TGTGCGTGAT	6780
ATTCGGAAC AATTCGTTT AACAGAAGCA GAAGAAAAGC AAATTCAGC TTTGATGAGA	6840
GAGCGGGGAG AGACTAATTT CTCTGATTTT CTTGTAATAA GTTTACTTTC CTCTGATTTA	6900
CAAAAACAGA TGGAGACATG GTTTGCCCTC TGGCAATCCC AAAAATAAGA ACAAATCAGT	6960
CGTGACGTTT ATGAAGTTT AATCTTGGA CAGTCAGAAC GTCAAGTCAC CCAAGAGCAT	7020
GTATCTATTC TCTTAACGTG CGTGCAGGAA TTGATTCAAG AGGTGCAAA CACCATACCC	7080
CTCAGTAAAG AATTCGTGA GAAGTACATG AGGTAAGCAC ATGGAACATC GTTACCGAAC	7140
CAATCTCAAG AAAGTGTTT TGTCTGATAG TGAGTTGAAC CAACTAAATA TAAATATCGA	7200
TCAAAGTGGT TGTAAATCCT TTTCTGAATA TGCAGACGA ACTCTACTCG ATCCTGGTAT	7260
GAATTTTATC ACGATTGACA CAAACGGTTA CCAAGATTTA GTGTTGAGT TAAAGAGGAT	7320
TGGCAATAAT ATCAACCAGA TTGCTCGAAG TGTTAATCAA TCTCAGTTAA TTTCTGGTGA	7380
AGAATTGCAG GAGTGAAAA AAGGAATTGG TGAATTGATA AAAGAAGTTG ATAAGGAATT	7440

TAATCTGCAA GCGCAGAAGC TAAAGGAGTT CCATGGTCAT CACTAAACAC TTTGCCATTTC	7500
ACGGAAAGAG TTACCGCAGA AAGCTTATCA AGTACATTCT CAATCCTGAG AAAACCAATA	7560
ATCTTGCCCTT GGTGTCGGAC TATGGCATGA AGAATTTTCT GGACTTTCCT AGCTATGAGG	7620
AAATGGTGCA GATGTATCAT GAAAAATTCA TCAGCAACGA TACGCTTTAC GATTTTCGCC	7680
ACGACAGGAT GGAAGAAAAT CAACGAAAAA TACACGCTCA CCACATCATF CAGTCTTTCT	7740
CGCCAGAGGA TCATATCACT CCTGAACAAA TCAATCGGAT AGGTTATGAG ACTGTGAAGG	7800
AATTAECTGG TGGCAAATTT CGTTTTATCG TTGCGACCCA TGTTGATAAA GACCACCTGC	7860
ACAATCACAT CATTATCAAT TCAGTAGATA GCAATTCTGA CAAAAAGCTC AAGTGGGACT	7920
ACAAGGTGGA GCGAAATCTT CGCATGATTT CTGACCGTTT TTCTAAAATC GCAGGTGCTA	7980
AAATCATTGA GAACCGCTAT TCTCACCAGC GGTATGAAGT CTATCGTAAG ACTAATCACA	8040
AGTATGAACT CAAGCAGCGA CTCTATTTTT TGATGGAACA TTCTAGGGAC TTTGAGGATT	8100
TCAAAAAGAA TGCTCCGCTA CTACATGTGG AGATGGATTT CCGTCACAAG CATGCCACCT	8160
TTTTTATTAC GGACTCAACT ATGAAACAGG TGGTGCCTGG CAAGCAACTC AATCGCAAGC	8220
AGCCTTACAC AGAAGAATTT TTTAAGAACT ACTTTGCCAA AAGAGAAATA GAAAGTCTCA	8280
TGGAATTTTT ATTGCTGAAA GTTGAGAATA TGGATGATTT ACTTCAGAAA GCAAACTTT	8340
TTGGACTAAC TATCAATCCT AAACAAAAGC ATGTTTCTTT TCAATTTGCA GGAGTGGAGG	8400
TAAAGGAGAC AGAGCTAGAC CAGAAAAATC TTTATGATGT AGAGTTTTC CAAGATTATT	8460
TTAAAAATAG AAAAGATTGG CAAGCTCCAG AACTGAGGA TTTCGTTCAA CTTTATCAAG	8520
AAGAAAAGTT ATCCAAAGAA AAAGAACTTC CAAGCGATGA GAAGTTCTGG GAGTCCTATC	8580
AAGAGTTCAA GAGTAACAGA GATGCCGTTT ATGAATTTGA GGTGGAGTTG TCACTCAATC	8640
AAATTGAAAA AGTAGTGGAT GATGGAATTT ACGTCAAGGT CAAGTTTGGT ATTCGTCAGG	8700
AGGGACTTAT CTTTGTGCCG AACATGCAGC TTGATATGGA AGAGGATAAG GTGAAGGTTT	8760
TCATCAGGGA AACCAGCTCC TACTATGTCT ACCACAAAGA CGCTGCCGAG AAAAATTGTT	8820
ATATGAAAGG TCGAACCTTA ATTAGACAGT TCAGCTATGA AAATCAAACC ATTCCATTAC	8880
GCAGAAAAGC GACAGTCGAT ATGATTAAAG AGAAGATTGC GGAAGTGGAT GCTTTGATTG	8940
AACTGGAAGT AGAAAATCAA TCTTATGTCA CGATTAAAGA TGAGTTAGTG CATGAACTAG	9000
CAGCGTCTGA ATTGAGAATC AATGAGTTGC AAGAACGAAT GTCAACCTTG AATCAAGTAG	9060
CAGAATATCT ACTGGCTTCA GTTGAAAGTA AGCAAGAAAT GAAATTAAAT CTTTCAAAAC	9120
TGAATATAAC TGAGAATATC AGTGCTAATA TTGTTGAGAA AAAATTGAAG AGCCTGGGGA	9180

520
ATCAACTGGA ATTGGAAAGG GGCAGGTATG AAAAGATGGT AGT 9223

(2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6827 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TCTGCTGGCT ACCATCATCT GACTTGGGCA AGACCAAAGT CTTAGTTACA ACTGTATTCT	60
TCTCAGCATT TTCAATAACT GGCAATGCCG ACTGAAGCGT ATCTTTTCT GTTTTTGTAG	120
CTGGTCCAGT TTCTTTTTC TGTCCGCAAC CAACCAGGAC AAAAAGGAAA GCTAGACTAA	180
CAAGAACTAT TTTTTCATT TCTTCTTCT TTCTTTTGA AATTAAAATA GAATAAGACT	240
GGGAAGTGCT CCCAGCCTTG ATGTTTATAG AGCTGCACGC AAACGTGCTT CTGCATTTTC	300
TACATTACGG ACAGAGCGTG GTAGGAAGGC ACGAATATCG TCTTCCTTGT AGCCAACTTG	360
CAGGCGTTTT TCATCTACAA GGATTGGGCT CTTTAAAATT CTCGGTGTTT CCATAATCAG	420
ATTGAGAACT TCATTGACAC TCAAATCTTC AATATCCACT CCAAGGGCTT TGGCATAGCG	480
ATTTTTAGAC GAAACGATGC TGGCTATTCC GTTATCTGTT TTGGTTAGAA TATCCAGTAA	540
TTCTTCTCTC GTAATTCCTT CTTTACCAAG GTTTTGTCTT TTATAACTTA ACTGGTGGGC	600
ATTGAGCCAG GTTTTGTCTT TTTTACAGCT AGTACAACCT GAGACTGTAT AAATTTTAAT	660
CATGTACCTA CCCCTTTCGC TACATGTTAC TATCAGTTTA GTCTATTATA CCATAAAAAA	720
CATCCGACTT GCGACCTATT TTTAATTTTT TTTGACTTTT TTCGTCATTT TCGTACTTTT	780
TTCTTGACAA ACAACTAAAT GACTATCAAC TCTTTTGGAG CTAGGGTCAA TAATTCACAA	840
CCTGTCTCTG TAATCAGGAT ATCATCCTCG ATACGAACGC CATATTTGCC TTCGATATAG	900
ATACCTGGTT CATCGGTCAA GGCCATACCT GTCTTAATAG TTTCTGTAGA AGTCTGACTA	960
AAGTAGGGTT CCTCATGGAT ATCCAGACCA ATACCGTGGC CAATGCCGTG AGTAAAGTAG	1020
TCACCATAAC CTGCCTCAAT GATAATATCA CGAGGGATTT TGTCAAAGTC ACGGAAACCT	1080
AAGCCTGCCT TAGCTTGGTC AATCAAGGCT TGGTTAGCTT TTAGAACCGT ATTGTAAATC	1140
TCTGCCTGCT CATCGCTAAC ATGCCCTAGA TAGATAGTCC GGGTCATATC ACTGACATAG	1200
TGGTCATAGA GACAGCCGAA GTCCATGGTG ATGGCTTCTC CCAACTCCAC TG GTTTGTGC	1260
ATTGGATGGG CATGGGGTTT AGAAGAATTG ATACCGCTAG CTAGGATCGT ATCAAAAGAT	1320
AAGCCAGATG CTCCCAACTC ACGCATGCGG AAATCAAGGA AGTTGGCAAT CTCAATTTCA	1380

GTTCCTCTG GTTTGATAAA GTCAAGCGCA TCGCGGAAAG CTTGGTCTGA GATAGAACAA 1440
GCCTTGCGAA TCGCTGCAAT CTCTGCCTCA TCCTTAATCA TACGAAGACC TTCCACAAAC 1500
TGAGTTTGTG GAAGCAAGTT CAAACCTGCA AAAGCTGCCT GCATACGGTG GTAATAAGAC 1560
ACTGAAATCT CATCTTCAAA ACCGATACGA GTCAAGCCCA TGTCTTAAC AATTCCTGCA 1620
ATGACAGCCA ATTCATCACG ATCAGCCACA ATCTCAAAAC CACTGGTTTC TTGCTTAGCT 1680
GCGATGATAT AGCGAGAGTC TGCTACTAAG ACCTGACGGT CACGACTGAT AAAGACTGTT 1740
CCGTTTGAGC CCCAAAACC AGTCAAATAA TAGACGTTTT TAAGATTGTT GATGATGATA 1800
CCATCTAGTT CTTTTCTTG CATTTTAGCT AGAAATGCTT GTACGCGTTT ATTCATGATG 1860
TAACTTTCCT TTCAAATAGT GTCCTGTATA GCTGGCTTCG TTGGCAGCTA CTTCTTCTGG 1920
AGTTCTCTGTT ACGATGATGG TTCCACCACC GACACCGCCC TCAGGTCCCA AGTCAATGAT 1980
ATGGTCTGCC GTCTTGATAA CATCCAGATT GTGCTCGATG ACGAGGACTG TATTGCCATC 2040
GTCTACAAAG CGAGCTAAAA CCTTGAGCAG GCGAGCAATG TCCTCTGTAT GAAGCCCTGT 2100
CGTCGGCTCA TCCAGAATGT AGAAAGATT TCTCTGTCGAT CGTTTGTGGA GTTCGCTAGC 2160
TAACTTCATA CGTTGGGCTT CTCCCCAGA AAGGGTGGTA GCTGGCTGTC CCAAGGTCAC 2220
ATAGCCTAGC CCTACATCCT TGATGGTCTG GAGTTTTCGT TGAATTTTCG GAATGTGTTG 2280
GAAAAATCTT ACCGCATCGT TGACCGTCAT ATCCAAGACC TGCAGAAATAT TCTTTTCCTT 2340
GTAGTGAAT TCTAGGGTTT CACTGTTATA GCGGGTTCGG TGGCAAACTT CACAAGCCAC 2400
ATAAACATCT GGCAAGAAGT GCATCTCAAT CTTGATAATC CCGTCACCTG AGCAAGCTTC 2460
ACAGCGACCT CCTTGACGT TGAAACTGAA GCGCCCCCTC TTGTAGCCTC GAATCTTGGC 2520
TTCATTTGTC TGAGCAAAAA GGTACAGTAT ATCGTCAAAA ACTCCTGTAT AGGTAGCTGG 2580
GTAGACCTC GCGTCCGTC CGATAGGGCT CTGGTCAATA TCAATCAAAC GGTCGACATG 2640
CTCAATCCCT GTAATAGTCT TAACTTACC AGGTTTGTCT GAATTACGGT TGAGCTTCTG 2700
GGCAATGGCT TTTTGAGAA TGCTGTTGAT TAGAGTCGAT TTCCCTGAAC CCGACACACC 2760
TGTCACGCG ATAAATTTTC CTAGTGAAAA GCGAGCCGTG ACATTTTGCA AGTTGTTCTC 2820
ACGCGCTCCT ATCACTTCAA TAAAACGACC ATTTCCGACA CCGCGCTCTT CTGGTACTGG 2880
GATGACACGT TTGCCTGACA AGTACTGACC TGTGATAGAC TTGCTGTTGC GAGCCACTTG 2940
CTTAGGTGTA CCTGCTGCAA CAATCTCACC ACCAAAAACA CCGGCACCAG GACCAACGTC 3000
AATCAGATAA TCAGCCTCAC GCATGGTATC TTCGTCGTGT TCCACCACGA TAAGAGTATT 3060
GCCCAAGTCA CGCATTTTT TCAGACTGGC AATCAGGCGA TCATTGTCCC TCTGGTGAAG 3120

522

ACCGATTGAC GGCTCGTCTA GGATATAGAG GACACCTGAT AGGTTGGAAC CAATCTGGGT	3180
TGCCAAACGA ATGCGCTGAC TTTCCCCACC TGAAAGGGTT CCTGCTGAAC GTGACAGGGT	3240
TAGATAGTTA AGACCCACAT TATTAAGGAA GGTCAAACGA TCCTTGATTT CCTTGAGAAT	3300
GGGACGAGCA ATGA'GGCTT CATTTTCAGA CAAAGTTAAC TGGCTCACCA AGTCCAAGTG	3360
GTCAGCGATA GACAGGTCTG AGATTCTCC AATATGTGGC CCTTGCTGGC CGCCACACG	3420
GACAGACAAG GCCTGGTCAT TGAGACGATA GCCTTGACAG GTTCCGCAGG TCAGCTCATT	3480
CATGTAGAGA CGCATCTGAG TGCAGTGTA ATCGCTATTG GTTTCATGGT AACGACGTTT	3540
GATATTATTG ATAACCCCT CAAACGGAAT GTCGATATCG CGCACGCCAC CAAATTCATT	3600
CTCATAGTGG AAATGGAATT CCTTACCATC TGACCCATAG AGAATCAAGT TCTTATCTTC	3660
TTCTGACAGG TCCTCAAAAG GCTTATCCAT AGCCACTCCA AAGACTTTCA TGGCCTGCTC	3720
TAACATGTTT GGATAGTAGT TGGATGAGAT AGGATTCCAA GGTGCTAGCG CTCCTCACG	3780
TAAGGTTTGT CTAGCATCTG GCACTACCAA ATCAGTATCC ACCTCCAGCT TGATGCCCAA	3840
GCCGTCACAC TCACTACAAG AGCCAAAAGG AGCATTGAAA GAAAAGAGAC GAGGCTCTAA	3900
CTCTGGGACA GTAAAACCAC AAAGTGGACA GGCATAATGC TCAGAGAACA ACAACTCCGA	3960
GTCGTCCATG GTGTCGATAA TGACATAACC TTCTGCAATA CGAAGGGCAG CCTCAATGGA	4020
ATCAAAGAGA CGACTACGAA TGCCCTCCTT GATAACAATA CGGTCAACCA CGACATCGAT	4080
ATTGTGTTGC TTGCTCTTAG ACAACTCTGG CACTTCGGTC ACATCATAGA CTTCCCCATC	4140
CACACGGACA CGAACATACC CGTCTTCTG AACCTTCTCG ATAACACTCT TATGTTGGCC	4200
TTTTTTCTTG CGGATGACAG GAGCCAAGAT CTGCAAGCGC TGGCGTTCAG GTAACCTCAA	4260
AACCTTATCA ACGATTGCT CCACAGAAGA AGCATTGATA GCTCCATGTC CGTTGATACA	4320
GTAAGGCGTC CCCACACGTG CGTAGAGGAG ACGCAGATAG TCATTGATTT CAGTCGTCGT	4380
TCCCACCGTC GAGCGAGGAT TTTTACTAGT CGTTTCTGG TCGATGGAAA TAGCTGGGCT	4440
GAGACCATCA ATGGCATCTA CATCTGGTTT TTCCATATTT CCCAAGAACT GACGAGCGTA	4500
GGCGGACAAA CTCTCTACAT AGCGACGTTG TCCCTCCGCA TAGAGAGTAT CAAAAGCCAG	4560
ACTGGACTTC CCTGAACCTG ACAAGCCAGT CACGACAACC AACTTGTCTC GCGGAATCTC	4620
CACATCAATA TTTTTTAAAT TATGGGCACG CGCCCCATGA ATGACAATTT TATCTTGCAT	4680
CTTTGTTCTT TCTAGTCCAT TATTGCTTAC CATTATACCA AAAAAAGTGA GATTCTATTA	4740
CCCAAAGGC CGATTTTGTA GTATAATAGT ACAGTGTGAA AAAATCTGAA AAATGAGAAA	4800
GGATAAGGGA TATGAAACAA GTTTTCTCT CTACAACAAC TGAATTTAAA GAGATCGATA	4860
CGCTTGAACC GGGTACTTGG ATCAATCTCG TCAATCCGAC TCAAAATGAA TCACTCGAAA	4920

523

TCGCCAACAC CTTCGATATT GATATTGCTG ACCTTCGAGC ACCGCTCGAT GCGGAAGAAA	4980
TGTCTCGTAT TACCATTGAA GACGAGTATA CCCTGATTAT CGTAGACGTG CCGGTCACGG	5040
AGGAAAGAAA TAACCGCACC TACTACGTAA CCATCCCGCT TGGTATTATC ATCACTGAGG	5100
AAACCATTTAT CACTACGTGT TTGGAACCAC TACCTGTCCT TGATGTCTTT ATCAACCGTC	5160
GATTGCGTAA TTTCTATACC TTCATGCGTT CACGTTTAT CTTTCAAAT CTTTATCGCA	5220
ATGCAGAGCT TTACCTAACA GCCCTTCGTT CAATCGACCG CAAGAGTGAA CAAATCGAAA	5280
GTCAACTGCA TCAATCAACT CGTAATGAAG AATTGATTGA GTCATGGAA TTGGAAAAA	5340
CTATCGTCTA TTTCAAGGCC TCCCTCAAAA CAAATGAGCG CGTGATTAAG AAATTGACCA	5400
GTTCAACCAG CAATATCAAG AAATACCTTG AGGACGAAGA CCTGCTTGAA GACACCTGA	5460
TTGAAACCCA ACAGGCCATC GAGATGGCAG ATATTTATGG AAACGTCTTG CATTCATGA	5520
CAGAGACCTT TGCCTCTATC ATTTCTAACA ACCAGAACAA CATCATGAAA ACCTTGCCCC	5580
TTGTGACCAT CGTCATGTCC ATCCCAACCA TGGTCTTTTC TGCCTACGGG ATGAACTTTA	5640
AGGATAATGA AATCCCCCTA AACGGAGAGC CAAATGCCTT CTGGTTAATC GTCTTTATCG	5700
CCTTTGCTAT GAGTGTCTCG CTCACTCTCT ATCTCATCCA TAAAAATGG TTCTAAGAGG	5760
AGTTCCTATG TCTCAAATG ATCTACAAAA ATAACTAAG AAAAACCAAG AGTTTGTCCTA	5820
CATTGCTACC CAACAATTCA TCAAAGATGG GAAAACAGAC GCTGAAATCC AGACTATTTT	5880
TGAGGAAGTC ATTCCCCAAA TCCTTGAGGA GCAATCTAAA GGTACAACGT CCCGTTCCCT	5940
ATACGGCGCA CCAACTCATT GGGCTCATAG CTTCAGTGC AAAGAGCAGT ACGAAAAAGA	6000
GCATCCAAAA GAAATGATG ACCCAAACT GATGATTATG GACTCAGCTC TTTTCATCAC	6060
TAGCCTCTTT GCCCTGTCA GCGCCCTCAC AACCTTCTTT GCGGCAGACC AAGCTTTCGG	6120
CTATGGATTG ATTACTCTTC TATTAGTTGG ACTGGTTGGT GGATTGCTT TCTACTTGAT	6180
GTACTACTTT GTTTACCAAT ACTATGGACC AGATATGGAT CGCAGTCAAC GTCCACCTTT	6240
CTGGAATCT GACTAGTTA TCCTAGCTTC TATGTTCTT TGGTTGCTTG TCTTCTTTGC	6300
AACAAGCTTC CTACCAGCTA GCCTTAACCC AGTACTGGAT CCATTGCCAC TAGCTATTAT	6360
TGGAGCAGCC CTCCTAGCCC TTCGTTCTA TCTCAAGAAA CGCTTGAATA TCCGTAGTGC	6420
AAGTGCAGGA CCAACACGCT ATCAAGAATA AGAAAACGAT AAAAGCAACT GCAGGTGCGG	6480
TTGCTTTTTC ACTTACTTTT TTGAGTTATA TTCAATGAAA ATCAAAGAGC AAACCTAGGAA	6540
GCTAGCTGCA GGTGCTCAA AGCACAGCTT TGAGGTGCA GATAAACTG ACGTGGTTTC	6600
AAGAGATTTT CGAAGAGTAT TAAAAGTATT CTCTCGAAAT CCCACATAGC TTTCTCTTAT	6660

524
ATTTTGTGAT AAAATAGGCT CAATCTATTT CTAGGAGGAT GAGATATGGT TTCTACTATT 6720
GGTATTGTTA GTTTATCTAG TGGCATTATC GGAGAGGATT TTGTCAAACA CGAAGTGGAC 6780
TTGGGTATCC AACGTCTCAA GGATCTGGGA CTCAATCCCA TCTTTTT 6827

(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11864 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CTGGCTAGTT GCATAGAGCA AAGTTGCTTC TTCATCAACA AAACCGTTCA TTTCAAAATA 60
GGAAAGCAGC TCATCAGGAC TCTCCAAACG AATCCCTTTG TAATCCAGCT CAACTGCCAC 120
CTCTTTCAAG GCTGCAAGAA GAAGTGTTCC CAGGCCCTGT CTCTGATGGT CAAACTCGAT 180
GACTAAAGAA TGTACTTTTA GACATTGCGG ATTGTCTGAC TGGGGACTTG ATAAAATATA 240
GCCTAAAAGT TGATTTTCAT CCCTAGCTAG AAGAAAGGTA TCCGCACACT TACGGATACT 300
TTCTTCTAAA ATATGGGAAA GTTGCTGCTT TTCAGCTGGA AAAGACGAGG TCTGAAGTGC 360
CCCTATCTCA GGCAAATCAG ACTTGCTTGC CTGAATGATC TTAATTGGAA TTTCCATGGG 420
AACATCCTAT TGAACATTGC TTGTCAAGTT AGACAAGAGA CGCTCAAATG AGTATTTCATA 480
GGTTTGGATG TCTCCTGCTC CCATAAGAC GTAAACAGCA TTGTCATGGT CTAGGAGTGG 540
AGAAACATTT TCAACAGTAA TCACTTGGTG TTTTGTGTG ATTTTGTGG CTAGGTCTTC 600
TACCTTAACG TCACCATGAT CTACTTCACG AGCCGAGCCA TAAATTTGCG CTAGATAAAC 660
AGCATCTGCT TGGTTTAAAG CATGGGCAAA GTCGTCCAAC AAGGCAATGG TTCTTGTAAG 720
GGTATGCGGT TGAAAGACTG CTACAATTTC CTTGCTTGGG TATTTCTGAC GAGCCGCATC 780
CAAGGTCCGA ATAATTTCTG TTGGATGGTG GGCAAAGTCA TCGATAATCA CTGTATCATTT 840
GACAATTTTC TCAGTGAAAC GACGTTTAAAC ACCGGCAAAT GTTTTCAAGT GCTCACGCAC 900
CAAGTTCAAA TCAAATCCTG CTGTGTAAAG AAGACCAATA ACGGCTGTCG CATTTCATGAT 960
ATTGTGACGA CCAAAGGTTG GAATGTGGAA TTGCCCCAAG TTTTGTCCAC GGAAATGAAC 1020
GGTGAAGGTT GAACCAGTTA TTGAACGAAG AAGATCACTA GCTACAAAGT CATTGCCTTC 1080
AGCTTCAAAA CCATAATAAT AAATTGGTGC ATCAGACGTA ATCTTACGCA ATTCAGCATC 1140
TTCACCATAG ACAAAAAGAC CCTTGGTGAT TTGTTTGGCA TAGTCGTTAA AGGCATTAAA 1200
AACATCCTCG AGACTTGTGA AATAATCTGG ATGGTCAAAG TCAATGTTGG TGATAATAGA 1260

525

GTATTCCTGGG TGGTAAGGCA TGAAGTGACG CTCATATTCG TCAGATTCAA AGACAAAATA	1320
TTTGGCATTG GCCGAACCAC GACCTGTCCC ATCTCCAATC AAGAAGCTGG TATCTGTAAT	1380
GTGAGACAAG ACATGAGACA ACATACCTGT CGTTGAAGTT TTTCCATGTG CTCCTGCTAC	1440
TCCCATGCTA ACAAAGTCAC GCATAAAGCT ACCTAGAAAC TCATGGTAAC GTTTGTAGCT	1500
GATACCATTT TGGTCCGCAT AGGCAATTTC GACGTTGTTA TCTGGACGAA AGGCATTTCC	1560
AGCGATAATT TCCATATCAC CGTCTAGATT TTTTTCATCA AAAGGAAGAA TGGTAATTCC	1620
TGCCTGCTCA AGACCGCGTT GGGTAAAGTA GTACTTTTCA ACATCTGATC CCTGAACCTT	1680
GTGCCCCATC TGGTGAACA TCAAGGCCAA GGCACCTATC CCTGATCCCT TAATTCCGAT	1740
AAAATGATAT GTCTTTGACA TGTTTTCTCC CCTATTCTGT CATTCTGGTC AGATTCAACT	1800
CTTGGGCAAC CCGACGTTCT TGTCTGTGTT GTTTACTTTT TTTATGTAG ATTTGGCTCT	1860
TCTTTAGAAA ATCATAATTG TTTTCTTTG GAGCAGGTGC TGACACTTCT TCATTCTTGG	1920
TAGGGATAGA ATGAACTTCT TCCGCCAAGA TATAATGAGA CTGGGTCAAT TTTTGGCTAT	1980
ATTTGACAAA TTCACCAGGA TTTTCTTTT GGAAAGGAGC TGTCGGTTGA TTGCCCTGTC	2040
TAACTAGACT GGGCTGAGAA TGACGTCTCG CAAGGCTGAA ATCCTGAGTT AGGTAGTTAG	2100
CAGAGCGTT CTTTTTCAAG TCCGCACGCG CTTCTTCACG CGCCACCTCC GCATAGCTCT	2160
TTCTTCTTTT TTTAACCCTT AAAGGAGCCT TTTTAGGTTT TTCGACTTGC TTTTCAATCG	2220
GTTTTACTGG TTTTCTTCA GCAATAGGAG CCCATTCTAA ATAATTTTTA TCTCGATACT	2280
CACCTTGAT ATTACTGATC AGATCAGACT CATCATAGAG ATTCATGACT GGCATTTACG	2340
TCAACATGAC CTCGTCATCT GACACCAATG GAAATCGTTC TTGTTTCATT TTCTATTTCC	2400
TTTCAACACT TCATTATAGC GTATTGTCTT GATTTTCAA GTGCTGGCTT CAGAAATTCC	2460
CAAAATTTCT CTAATTTCTG CTAGGGTCAG ACTACCAGT GACTCTGTGC CGTCCAATAC	2520
TTGTGACACC AGATGTTTCT TTTGTTCTTG GAGTTCCTGA ATTTTTTCTT CAATGGTTCC	2580
CTTGGTCACC AAGCGATAGA CCTCAACCGT TTCTTCCTGA CCCATCCGAT GGGCACGGCC	2640
AATGGCTTGC GCTTCCACCG CAGGATTCCA CCAAGGTCA ACCAAGATCA CTGTATCTGC	2700
ACCTGTCAGG TTCAGACCGA CCCCAACAGC CTTGAGGGAA ATCAGAAAGG CATCTCTTTC	2760
TCCTTGGTTA AAGGCCTTGG TCATGTCTTG TCTTTCCTTG GCTGGGGTTG AACCCGTAAT	2820
TTTAAAGGAA GTCAGGCCCA AGTCTGGCAG TTCTTGTTCA ATTTTTTCCA ACATTCCTT	2880
GAACTGAGAG AAAATCAAGA CACGGTGTCC GCCGCTGCC ACCTGTACCA GTAGGTCTCG	2940
GAGACTATCT AGTTTGCCGC TGGCTCCCTG ATAATCTTCC ATAAACAGGG CAGGAGTGTC	3000

526

ACATATTTGA CGCAAGCGCA TCAAACCAGA TAAATTTCC ACACGACTTC GCTGAAATTC	3060
CTGTTCTGAC ACTTGAGCCA GATGGTCTCG CATCTGTTGT AACTGGGCAA GGTAATAGC	3120
CTTTTGCTGG TCTTCCAGTT CATTTTATA AACCACCTCA ATCAAGTCTG GCAATTCAGT	3180
CAGAACTTCT TCTTCTTGC GTCGCATCAC GAAAGGCTTG ATAACTGAG CCACTCGCTC	3240
TGCTGGCAAT TTCATAAATT CTTTCTTGCT TGGCAAAAGT CCAGGCATGA CGATTTGGA	3300
AATAGACCAC AACTCACCCA GATGGTTTTC AATCGGAGTT CCTGACAAGG CAAAGACCGA	3360
CGGCACCACA AATTGTCTCA AGGTCTGGGC AATCTTGGTC TGGGCATTTT TCATGACCTG	3420
AGCCTCATCT AAGAAAAGGA AGTCAAAGGC CATCCCTTGA TAAACTCAC TGTCTGACG	3480
GAAGGTGGCA TAGCTAGTCA CATAGATTG ATGGCTCTCG GCAAGAATCT CCTCACGACT	3540
TGCTTTCAAA CCATGAACAA CAGTCACATC CAACTGTGGA GCAATTTCT GAAACTCATC	3600
TGCCCAGTTG TAAATCAAAC CCGACGGAGC GAGAATCAA ACCCGACTTT CTTTGTCTAC	3660
TTGACTAGTC AAAAAAGCAA TGGTCTGAAG GGTTTTCCCA AGTCCCATAT CATCAGCCAA	3720
AATCCCACCA AAACCATAAT GATGGAGCAT CTGCAACCAG CCAATTCCTT TTTCTGATA	3780
ATCTCGCAAG TCAGCCTTGA CTTGAGTTGC TTGCAAAGGA AAGTCTCTG GATGCGTCAA	3840
ATCCTGGGCC AGATTCTGGA ATTCTTGTGA AAAAGAAACA CGGTCTCGCC CTTCAAAGAG	3900
ATGAGCTAAA CTGTAGGCCA AGGATTTCGG AGCCTGCAAG GTCCCATCTT TTAATTCAAA	3960
TGCCCCAGT TCCTGTAGAT TTTGGCGAAT TTTCTTGGT TCTTCATCGA AAAAGTAAAC	4020
TTGATTAGAC GAATCAATAT AAAATCCTG ATTGGCAACC AAGGCCTGCA TGGCTTGGTC	4080
GATTTCTTCC TGGACAATAT TTTGAAAATC AACTGGATT TCCAAGAGAC CTCCCTTGGA	4140
GGCAATCTGC ACCTGAGGAC TCGCTAGGCT ATAAAGCTCT TCTAGTTTAT CTGATAGGTC	4200
AACATGCCCG AGTTTTCCTA AGACTGGAAT GATATCATGA AAAAAATGAT AGACAGACTC	4260
CGCTTTTAAG GCCTGACGCC AAGATTGAAA ATCGGCCTCA AAGCCCGCAG CCAAACAGAC	4320
TTGGAAAATT CTTTCTTCTA AGTCTGCGTC ACTTGAAAAG GGTAAATCTT CTAGCTCTTG	4380
TCGGCTAGAT ACCTGTCTAT TTCCATAATC AACTGGAAT TCTAAACGAA TCCGATTATC	4440
TTCTTCCCTG TCAAAGTAAA AAGAGGGCGC AAAAGTTTG ATTGTAGAC GTTCTGGAGC	4500
TGAAACGGTG CCCATCTGGA TAAAAAGAGT CAGACAGGAG GCCAATTGT CTGATCACT	4560
GCTATCAAAT TGCAGTATT TCTTCTCTTG TTGACCCACA GGTAACGCTT TAATTTCTT	4620
GAGAAGACGC ATCTGCTGGT CTGTAAAAA ATAAACCTGA CCTTTATGGA AAAGTACTGC	4680
TCCCTGATAA AAGACATTGA CCTAGGACT CTCCTGATT TCCATTTCOA AATAATCCGA	4740
GTATTCTGTT ACTGTAAAGG CAAATAGATT GGCATCAGCA TGCATATCCT GAAAAAGCAG	4800

527

GGTTTGGTAG CTATCCACTT GATGGTCAAA TTGAAAATGG GGCAAGGCCA TCAGTAAATT	4860
CACACCCTGC TCAAAAAGG TCAGAGGGAA AAAGAGGTGC CGACCTTGGT TTTGGAAAAA	4920
GAGGTCTGGA ACCAGCCCTT CCTCCGTTAG TCCGTGCAAG AAAGTCAAAA GTTCTTGGCT	4980
GGCATCATCA AAGGCTTCCC AAGAAAGAGA CTCCTCATAA ATCTTGCCAA TCATATACGA	5040
CTTCTCTGTC TCGACAATCC TTAAAAAAG TGGAATATCC CGAATGACAT AGTATTTTGT	5100
GCTATTGATT TGGCCGATTC TCAGAGTCCA CAAGATATGA TTGGTTCCTG CTTCCACCTG	5160
ACCCACAGCT GATAACTCAT AGGCGCATTC TGATTTTGA GATAAAATTC GATCCAAAAA	5220
CTTGCCACCC AAGGTCACCT TGGTTTCAAC AGCCTCTTTT TCTTCATGAC CTTCTTCCAG	5280
ACTCCACAAG ATTTCTGAC CACGCTCATC ATTTTTCAGA AAATGCTCTA GCGCTGCCAA	5340
ATGCACACAG TAGCCCTCT TTTGAAAAA ATCACAGGCA CAAAAACCA AATCATCCTC	5400
TAAACTATAG CGCAGTCTT CTTCTGCAAC GCGAGCGTAG AGCCGATTGT TCTTTTCTT	5460
GATGATATCA ACCTTACCAG TTTCATAAAG GGCAACACCT TCGATACGAA TTTTCCCCGG	5520
AATCAATTTA GCCATATTTT CACCTTTACC TTATCTTTT ATTATACCAT ATTTTCGCTT	5580
ATGAAAATAG CCTTCTAGGA AGACTTTTCT CCTAGAAGGC TGGATTTTTA ACGTTTGGCA	5640
AAAGTAGCCA CAATCCGCTG ACAGACTTCT TGCAACAGAG ATTTGGGCAT AGCTATATTG	5700
ATGCGGGCAT GGAGACTTCC TTCCTCTCCA AAATCCAAAC CACGGTTGAG GATAACCTTG	5760
GCTTCATTTT TCAACAACTC TTGCAATGTT TCATCAGTCA GGTCAATAGC TGAAAAGTCA	5820
AGCCAAATCA AGTAGGTACC TTGCGGTTTC ATGACCTTGA TTTTAGTCTC TTTTCCAAAT	5880
AGATCCATCA CATAATTGAT GTGGTCTTCA AAGACTTGCT TGAGTTCCTC TAGCCAATCT	5940
TTACCGTATC GATAGGCAGC TTCTGTGCGC AAATAACCCA AGCCTGAAAT TTCATGCTGA	6000
TTATTGGCCA ACAGGCGTTT CTGGAAGGCC AGTCTCAACT TAGGATTTTC AATGACTGCA	6060
TAGGAATTTT TTGTTCCAGC AATATTAAAT GTTTTAGTGG CACTGCTCAA GACGATAGCA	6120
AAATTTTGA AGGCAGGATT GATGGTATTG AAAGACTGGT GTTTGTGACC AAAGAGGGTC	6180
AAATCTTGGT GAATCTCATC CGAAACTAAC AAAACACCGT GTTTTGGCA GAGTTGGCCA	6240
ATCTTCTCCA AACTTCTTT TTCCCAAACA CGTCCACCAG GATTGTGAGG GTTGCAAAGA	6300
ACATAGAGTT TAACCTCCTC TTCCACCAA TCCTTTTCAA GTTGGTCAAA GTCAATCTCA	6360
AACAGACTAT CCTTTTCCAC TAAGGAATTA GTAATCAATC TACGATTATT CAACTTGACA	6420
CTGCGAGCAA AGGGTGGSTA GACAGGCGTG TTAATTAAAA CCGCCTCGCC TTCTTTTGTA	6480
AAGGTTTGAA TAGCTGTTGA GATGGCTGGT ACCACACCTT CGATAAAGAC AAGAGCCTCT	6540

528

TTGTCAAAGT TGTAACCGTA TTGTGTAGCT TCCCACTTTT GAACTTCCTT AATTAAGTCT	6600
TCACTGGCAT AGGTATAACC ATAAACCAGT TGGTCTGCGT AAGTTTGCAC GGCTTGGCGG	6660
ATTTCAAGCA AGACCACAAA GTCCATATCC GCTATCCAAG CTGGTAGAAC TTTACTATCC	6720
GTTTCTGTTT CTTTCCATTT ATAGGTATGG TGCCCTAAAC GGTGCGGAG GCTTGTAAAA	6780
TCATATTTTC CCATCTTTGT CTTATCCTTC TATGGCTTGG CGCAAATCTG CAATCAAATC	6840
TCTAGCATCC TCAATCCCAA TAGACAAACG CAAGAGGTCA TCTGTCAAAC CATAAGAATG	6900
GCGTACCTCT GCTGGAATAT CAGCATGAGT TTGAGTCGTT GGATAAGTAA TAAGACTTTC	6960
CACTCCACCC AAACCTTCCG CAAAAGAGAA GACCTTGAGA CTGTCAAAA TATGAGGAAT	7020
GCGTGTTC A TCGGCTACTT TAAAGGAAAT CATGCCCTCA CGACCAGTGT AGAGAACTTC	7080
CTTAACGTCT GGAGAATCCT TCAAAAAGGC AACCCTTCTT TGGGCGTTAG CTGTTGAGCG	7140
CTCCATACGA AGAGACAAGG TCTTGAGACC ACGAAGCAAC TGGTAGCTGT CAAATGGAGA	7200
CAAGACTGCC CCTGTTGTAT TAAGATTGTA AAAAAGCTTC TCGTATAGTT CTAAACTATT	7260
GGTCACAACC ACTCCAGCCA AGACATCATT GTGGCCTGCT AGATACTTGG TTGCTGAATG	7320
GAGAACGATA TCTGCTCCAT CTTCAATCGG ACGTTGGTAG ATAGGGCTAT AGAAGGTATT	7380
GTCCACCACC ACTTTGGCAC CCTTAGCATG AGCCAAATTTT GCTAGTTTTT CGATATCAAA	7440
TTCCAACATC AAGGGATTGG TTGGGTTTC GATATAGAGA ACATCCACAT CCTTTTCTAA	7500
CTCGGCAATC AACTCTTCTT CTGTATTGGC ATAGGTAAAA TGGAAATGAC CTTCTGCTC	7560
CACTTGGTTA AACCAGCGAA AAGAACCACC GTAAAGATCA CGCACTGCCA AGACCTTACT	7620
TCCTACTGGA AAGACGCTAA AGGCCAGTAC AATAGCTGAC ATCCCTGAGC TAGTCGCTAG	7680
GGCATAGTCT GCTGACTCAA TAGCCGCCAA GACTTCCTCA GCCTTACTAC GAGTTGGATT	7740
TTTAGTGC GC GTATAGTCAA ACCCAGTAGA TCGACCAAAC TCTGGATGCT GATAGGTCGT	7800
TGAAAAATGA AGTGGTGTC CCAAAGCACC TGTTCCTCA TCAGACTTGA TCCCTGCTTG	7860
TGCTAAATTT GTGTTAATGT GTAATTCCTT GCTCATACAA TTCCTCCAAA TCTATAGTAA	7920
CTATTGTACC ACTTATTTTG TATCCTTCGT TTTCTTGTTT TCAAGAGCTA GTTATAGTTT	7980
CAAATATAT AAAAAGGGAG TTTTTCCTGC TCCCTTTAAT AGACTATAAA ATGGTGAATC	8040
TCAAAAGACA CCTTCACTCT ATCATTGCT CCTGCACAAA ACGAGCATAA CGCTCATGAT	8100
TTTCCAGTAG TTCTTTATGA GTTCCTGAGC CAGTGATTTT CCCCTCCTCT AAGAAGAAAA	8160
TACAATCCAC ATCTTTTACC GTTGACAAAC GATGCGCTAT AATCACAACC GTCTTCTCCT	8220
TTAGTACAGA ATAGAGGCTA CTGATAATCG CATACTCAGA ATCCGCATCA AGATTAGCAG	8280
TGGCTTCATC AAATATAAGA ATTCAGCAT CTTTAAAGTA GGCTCTAGCT ATTTGAAGTC	8340

529

TTTCGPTCGC CCCCCTGACA AGAGTCGTCC GCGTTCACCA ACTTCAGTAT CTAGTCCCTC	8400
TTTCATGGAG CGAATCTCAT CACCTAGTGA TACTAAGTCT AGCACTTTCA TCAATTCATC	8460
ATCAGTTACT AAGCGATTCA AACCGAGACA AAGATTGTCA CGAATACTGC CAGATAAGAC	8520
TGCATTATTT TGTGAAACCC AAGCGATTTT ACTTCTCCAT TCTTTTAAGT TAAAAACATA	8580
TATACTTGAT TGCTCCATTA GAATATCTCC TGAAAGCGGT TTATAAAACC GCTCTAACAA	8640
ACGCACAATC GTTGATTTTC CTGATCCAGA TGGTCCAACA AAAGCAATTT TTTGCCCTT	8700
GAAAAATGAA CAAGTAATAT CCTTTAAGAC AGGTCGATTT TCATCATAAC CAAAATAGAC	8760
ATGGTTAAAA TTCAACCCCTC GTCCTGATAC CGATTTTCTT CCCTCAAATT TTTCTTTAGG	8820
AACTGCAAGC AAGTTCTCCA GTGCAACTGA AGATCCCTTG CTCCTAGAAT AAACAGTTAC	8880
AAAATTAGCT ATATTACTAA TAGGATTAAG TAATTGAAAG AGGTAAATCA AAAACGAAAC	8940
CAAGGTTCCC ACAGATATAT ATCCTGCGCT GACCCGATAA CCCCATAGG TTAGCATCAC	9000
AGCTATAGTC GCAAAGATAA ATAAGAGAGC AAACGGGGTC TCAAAAGAAG TAACCCTATC	9060
TGATTTCACT GAATTGTTTT GTACCCTTTC AATACAATTA TCCAAAACAT CCTGTACACT	9120
TTTCTCTGCT TGGTTAGTCT TAATTAATTC ATGTTCTTGA ATCTTTTCAG TCAATTGCCC	9180
TGTTAAATTT CCTCCTGTAA ACGACGACTA TACTTTTCAC TGATATTGGA AAGGGGCAAG	9240
ATAATAACA TCATACAAGG AAGAGTGATG AATAAAAGTA GAGAAAGATT CCAATCAAGA	9300
CTAAATAAGA CTACAATGGA ACCAAGTACC ATAACATAAC TCAGAATAAT ATTTGGGAAA	9360
GTCGTAATTA AAAACTCACG AATGACACTC GTGTCATTGA CAATGGCAGA AGTCAACTCC	9420
CCACTTTGGC TCTTATCAAA GAAGGATTTT TCTACATAAA TCAACCCCTC TATCACTTTT	9480
TTCTTGATTT TTGCTATCTT TTTTTCACCC GATTGACTAA ACAGATAGTA ACCAATAGAA	9540
GAAAACAAGG CTTGACCAAT AAAAATCAAA AACGATTGAA ATACTTTGGA GCCTATATTT	9600
TCAATAGAAC TCCCATCTAT TAAATCCTTT AAGATAAGGG GAAGCAACAA AGCAAGTAGA	9660
CTAGACAGAA CAAGTAAGAA ACTCCCCATA ATCACCTTAG TATCTACTCT TAATAATTTT	9720
AATTTCATAA ATACTCCTTA TAATATTTCA ACGGATAAAG TCGGGAATAA CTCAATTTGA	9780
GGATAAAATC TAATAAATCT TCCTATAACA AAACGCATAA CATCTAGGAT TTTATATACC	9840
TGATATTATG CGTTTTTAAG CACAAAGACT TCTTACACAA ACTTATCTAC AATTAGATTT	9900
TATTTGACAT GTTTTGCCAA TTCTTCTTGG GCTTTTTTAT TGGATTCTTC TTTTCTTTC	9960
AACCATTTTT CTCTGGCTTT TGCATATTCG TCTGTTGTGA CAATCTTATC TTGTACTTTG	10020
AGGTATTTAT ATGATTCAAC CCCTTTTGTA CCGGTTAAAC CATAGGCAGC AGCAAATGGT	10080

530

ACGGTTCCTC TCAATGATGG TGTTCCCCCA CGCGAAACAC TTGGAAGAAC TAAAGAACTA	10140
TCAATCAACC AAGCTTGAAT ATCAGCATAT TTCTCATAAC GTTTGGCCGG ATCTTGCTCT	10200
TTATTAGCTT CTTCCAACAT TTGAGTATAG ACATCCAGTC CAACTGCCTT AGCCTTGTC	10260
TTGGCCTCAC CAGGCTCTAG TCCAAGATT TGCAGAAATC CTCCACTATT AGTATTAAAA	10320
ATATCGAGAT AGGTTGACGG GTCTTGATAA TCAGGTCCCC AACCGCCATG ATATAAATCA	10380
TAATCTTTCT GAGCAGCTGT TTGAGCAAAG TAGCCTGAAC TGTCAAATC ATCTGATGTT	10440
AATTGCTGAA TGTCAATCAC TACATTATCA GAACCTAAAA CAGATTCAAT TGATTGTTTG	10500
ATAGAACTAA CTCCTTGAT GCCTACTTTA TCTGTTACTT CCACAGTCTT ATCCAAGTGG	10560
ATTGGGAATT GAACACCCTT TGCTTCGAGT TCTTTCTTAG CTTCGCCAAA CTTAGCCTTG	10620
GCTTTCTCAG GATTGTAGTA AGGTCTTGA CCATCCGCAA AGTTGATACC TTGCCATTCC	10680
TTACCATAGT TGACCATCTT AGAGGCTACA ACTTCACCAA AGTCTTTTCC CTTGATACTG	10740
ACAAAGTTTG GAGGAACCAC TAGGTTACGC AAAATCTTTG TTGCACCTTC TTTCCCTTCA	10800
GACTGAGCCC CATAAGATGT TCTGTCAAAA GCAAAATTGA TAGCCTGACG GAAGTTTTTA	10860
TTGAGAACTG CTTCTGAGT CGATTTCTTT TCAATGTCAC TTGTTTTAGA AGTATAATTG	10920
TAAGACTTCC TATCTAGGTT AAAATTAAAG AAATATGAAG TTGAATTTTG CATACTATAG	10980
ATGATATTGT TTTTGATTTT TTCTTTAATC CCTTCATAGC TGGAGCTGTT AGGAAAAAGA	11040
CGAGCCGTAG TATAAGCACC AGCTGTAAAA TTACGTTCCA GTGATTCTTG GTCGCTACCA	11100
TCATAGTAGG TCAATTTTAC ATCGTCTACA AAGACATTCT TAGCATCCCA GTAATTAGGG	11160
TTTTTCTTAT ATTCAATAGC AGATTTTGAG ACAAGTGCTT TCATCAAGAA AGGTCCATTG	11220
TACAAAATAC TAGATGGATC CGCCTTCCCA AAATCATCCC CTTTGTGATT CAGGAAATCT	11280
GCATTAACAG GAAAAAGTAT CGTTGCAAGT GTTTTGAAT TCCAGTAAAG TTCTGGTTTA	11340
ACCAAAGTAT ATTGAACCGT TTGGTCATCA AGTGCTTGA CACCGACAGT TGAAAAGTCG	11400
CTTGTTTTAC CAGTGATATA GTCATCCAAA CCAGCAACAG AGTCCTGCAC TAGATACAAG	11460
GCTTCTGATT TTTTATCAGC TGCATATTGC AAACCTGTCA CAAAATCCTG GGCAGTTACA	11520
GGCGCATATT CTTCTCCCTC AGAAGTAAAC CACTGGCAT CCTTACGAAG TTTGTAGGTA	11580
TAGGTCAAAC CGTCCTGAGA AACAGTCCAA TCCTCTGCTA ATGATGGAAT AATATTCCCA	11640
TATTGGTCAT TTTCTAATAA CCCGTCTACC AAATTGCAA CAATATCGGA TGTGCTGCG	11700
CGGTTTTCTG CTAGATAGTT CAAGCTAGAT GGATCACTG AATAAACATA GTTGTAGGTT	11760
TTTGACGCCG TGCTAGAATT TCCACACGCG CTCAATAAAA CTCCTGTACC CAGGACAAGA	11820
CCTGCCAAGG TTAGATATTT GCTCTTAGAC TTTTTCATTT CCGG	11864

531

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2412 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TAAGTGCCTT AAACATAATA TAAGGAGAGA AAATGTCTGC AATAGAACGT ATTACAAAAG	60
CTGCTCACTT AATTGATATG AACGATATTA TCCGTGAAGG GAATCCTACT CTACGCGCGA	120
TTGCTGAGGA AGTCACTTTC CCCCTATCTG ACCAGGAAAT CATCCTAGGC GAAAAGATGA	180
TGCAATTCCT TAAACATTCC CAAGATCCTG TCATGGCTGA AAAAATGGGA CTCGCGGGTG	240
GTGTTGGACT GGCTGCTCCC CAGTTAGATA TCTCAAAACG CATTATCGCT GTTTTGGTAC	300
CTAATATTGT TGAAGAAGGC GAAACTCCAC AGGAAGCCTA CGATTTGGAA GCCATTATGT	360
ACAATCCAAA AATCGTCTCT CACTCTGTTC AAGATGCTGC TCTTGGCGAA GGAGAAGGTT	420
GCCTGTCTGT TGACCGTAAC GTGCCTGGCT ATGTTGTTTC CCATGCCCGC GTTACTGTTG	480
ACTACTTTGA CAAAGATGGA GAAAAACACC GTATCAAACT CAAAGGCTAC AACTCCATTG	540
TTGTTTCAAG TGAAATTGAC CACATTAACG GTATCATGTT TTACGATCGC ATCAATGAAA	600
AAGACCCATT TGCAGTTAAA GATGGTTTAC TGATTTCTGA ATAAAGAAAA TCCCCTTGCA	660
AGACGGGGTT TTGTGTTATA ATAGAGGCAT GAAACAAAT GATATTGTCT ATGGTGTCCA	720
CGCCGTTACC GAAGCCCTCC TTGCAAATAC AGGAAACAAA CTCTACCTCC AAGAAGATCT	780
CCGAGGTAAG AATGTTGAGA AAGTCAAGGA ACTAGCTACA GAAAAGAAGG TGTCCATTTC	840
TTGGACATCA AAAAAATCTC TCTCTGAGAT TACTGAAGGT GCTGTTTCATC AAGGTTTGTG	900
TCTACGAGTG TCTGAATTTG CCTATAGCGA GCTAGATTAC ATCCTTGCAA AACACGCCA	960
AGAAGAAAAT CCACTTCTAT TGATTCTAGA TGGTCTAACC GATCCCCATA ATCTGGGTTT	1020
TATCTTGCAG ACAGCCGATG CGACCAATGT TTCAGGTGTC ATCATTCCCA AGCACCCTAC	1080
TGTCGGAGTA ACTCCTGTCG TTGCCAAAAC AGCCACAGGT GCTATTGAAC ACGTtCCAAT	1140
TGCCCCGAGT ACCAACCTCA GTCAAACCTT AGGATAAACT TAAGGATGAA GGTtTCTGGA	1200
CCTTTGGAAC GGATATGAAC GGTACTCCTT GCCACAAGTG GAATACAAAA GGGAAAATCG	1260
CCCTCATCAT TGGAATGAA GGAAAAGGTA TCTCTAGCAA CATCAAAAAA CAGGTCGATG	1320
AAATGATTAC CATTCCGATG AATGGACATG TTCAAAGCCT TAATGCCAGT GTTGTGCGG	1380

532

CCATTCTCAT GTACGAAGTT TTCCGAAATA GACTATAAAA AAGTTTCCAG TCATCTGATT 1440
GGAAACTTTT TTATGATTAA CTATGTTCTG TAATGAATTT ATAGGCTTCT TGACCAGCGA 1500
TAGCTCCATC TCCAACCGCT GTTGTACTT GGC GAAGGTC TTTCAAGCGA ACATCTCCAA 1560
CTGCAAAGAT ACCGTCGACT GCAGTTTCA TGTGGTTATC TGTCACAATC CATCCTGCCT 1620
GATCTTGGAT ATTCAATTCT TTAACAAAAT CGCTAAGAGG GTCCAAACCA ACATAGATAA 1680
AGACACCACC GAAGGCTTGT TCTGTCACTT GACCTGTTTT CACATTTTCA AATACGACTG 1740
ATTCTACTCG GTTTCACCC TTGATTTCCC TTAACACAGA ATCCCAGATA AAGCTGATTT 1800
TTTCATTCGC AAAGGCGCGA TCTTGTA AAA CCTTTTGGG AC GAAGTTGG TCACGACGGT 1860
GAACAATGGT AACAGTCTTA GCAAAACGAG TCAAGAAGAG GGCTTCTTCA ACAGCTGAAT 1920
CTCCACCACC AACTACCAAT AAATCTTGGT CACGGAAGAA AGCACCATCA CACACAGCAC 1980
AGTAAGAAAC ACCACGACTG TTCAGTTCTT CTTCTCCAGG CACTCCCAA GGACGGTGTT 2040
TAGAACCAGT TGCTACGATA ACTGTACGTG TTTTATATGT TTGGTCATCA GTCATCACTT 2100
TCTTAAATC ACCATGGCTT CGACATTTTC AACATAACCA TAAATGTGCT CAACACCAAG 2160
ATTTTCAAGT GGTTCAAACA TCTTTTCAGC CAATTCAGGT CCACTAATAT TAGCGTATCC 2220
TGGGTAATTT TCGATATCAG ATGTATTATT CATCTGACCA CCTGGCAGAC CACCTTCAAT 2280
CAAAGCTACT TTTAGATTGC TTCGAGCAGC ATACAAGGCC GCAGTCATCC CTGCAGGTCC 2340
AGCACCGATA ATAATAGTAT CGTACATATA GATTCTTCT TCTTGGTGT AACTATCTTT 2400
ATTCTAACTC TG 2412

(2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7760 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CCGATTTGGT GGAATTTTGG TCTCATCATT TAGAAGGTGT TGCAAGAGCA GAGTTTACCT 60
TGGTGCTTCA TACCAAATTG GGAGAAGCCT CTGTTTGGC AAATATTGTA GATGTAAACA 120
AGGATGAATG GATTTTAGGA ACAGTTGCTG GTGCCAATAC CTTATTGGTT ATTTGTCGAG 180
ATCAGCACGT TGCCAAACTC ATGGAAGATC GTTGTCTAGA TTTGATGAAA GATAAGTAAG 240
GTCTTGGGAG TTGCTCTCAA GACTTATTTT TGAAAAGGAG AGACAGAAAA TGGCGATAGA 300
AAAGTTATCA CCCGGCATGC AACAGTATGT GGATATTAAA AAGCAATATC CAGATGCTTT 360

533

TTTGCTCTTT CGGATGGGTG ATTTTATGA ATTATTTTAT GAGGATGCGG TCAATGCTGC	420
GCAGATTCTG GAAATTTTCT TAACGAGTCG CAACAAGAAT GCCGACAATC CGATCCCTAT	480
GGCGGGTGTT CCCTATCATT CTGCCCCAACA GTATATCGAT GTCTTGATTG AGCAGGGTTA	540
TAAGGTGGCT ATCGCAGAGC AGATGGAAGA TCCTAAACAA GCAGTTGGGG TTGTTAAACG	600
AGAGGTGTGT CAGGTCATTA CGCCAGGGAC AGTGGTCGAT AGCAGTAAGC CGGACAGTCA	660
GAATAATTTT TTGGTTTCCA TAGACCGCGA AGGCAATCAA TTTGGCCTAG CTTATATGGA	720
TTTGGTGACG GGTGACTTTT ATGTGACAGG TCTTTTGGAT TTCACGCTGG TTTGTGGGGA	780
AATCCGTAAC CTCAAGGCTC GAGAAGTGGT GTTGGGTAT GACTTGCTCTG AGGAAGAAGA	840
ACAAATCCTC AGCCGCCAGA TGAATCTGGT ACTCTCTTAT GAAAAAGAAA GCTTTGAAGA	900
CCTTCATTTA TTGGATTTGC GATTGGCAAC GGTGGAGCAA ACGGCATCTA GTAAGCTGCT	960
CCAGTATGTT CATCGGACTC AGATGAGGGA ATTGAACCAC CTCAAACCTG TTATCCGCTA	1020
CGAAATTAAG GATTTCTTGC AGATGGATTA TGCGACCAAG GCTAGTCTGG ATTTGGTTGA	1080
GAATGCTCGC TCAGGTAAGA AACAAGGCAG TCTTTTCTGG CTTTGGGATG AAACCAAAAC	1140
GGCTATGGGG ATGCGTCTCT TGCCTTCTTG GATTCATCGC CCCTTGATTG ATAAGGAACG	1200
AATCGTCCAA CGTCAAGAAG TAGTGCAGGT CTTTCTCGAC CATTTCTTTG AGCGTAGTGA	1260
CTTGACAGAC AGTCTCAAGG GTGTTTATGA CATGAGCGC TTGGCTAGTC GTGTTTCTTT	1320
TGGCAAAACC AATCCAAAGG ATCTCTTGCA GTTGGCGACT ACCTTGCTCTA GTGTGCCACG	1380
GATTCGTGCG ATTTTAGAAG GGATGGAGCA ACCTACTCTA GCCTATCTCA TCGCACAAC	1440
GGATGCAATC CCTGAGTTGG AGAGTTTGAT TAGCGCAGCG ATTGCTCCTG AAGCTCCTCA	1500
TGTGATTACA GATGGGGGAA TTATCCGGAC TGGATTTGAT GAGACTTTAG ACAAGTATCG	1560
TTGCGTTCTC AGAGAAGGGA CTAGCTGGAT TGCTGAGATT GAGGCTAAGG AGCGAGAAAA	1620
CTCTGGTATC AGCACGCTCA AGATTGACTA CAATAAAAAG GATGGCTACT ATTTTCATGT	1680
GACCAATTCG CAACTAGGAA ATGTGCCAGC TCACTTTTTC CGCAAGGCCA CGCTGAAAAA	1740
CTCAGAACGC TTTGGAACCG AAGAATTAGC CCGTATCGAG GGAGATATGC TTGAGCGCG	1800
TGAGAAGTCA GCCAACCTCG AATACGAAAT ATTTATGCGC ATTCGTGAAG AGGTCGGCAA	1860
GTACATCCAG CGTTTACAAG CTCTAGCCCA AGGAATGCG ACGGTGATG TCTTACAGAG	1920
TCTGGCGGTT GTGGCTGAAA CCCAGCATTT GATTCGACCT GAGTTTGGTG ACGATTACACA	1980
AATTGATATC CGGAAAGGGC GCCATGCTGT CGTTGAAAAG GTTATGGGGG CTCAGACCTA	2040
TATTCCAAAT ACGATTGAGA TGGCAGAAGA TACCAGTATT CAACTGGTTA CAGGGCCAAA	2100

534

CATGAGTGGG AAGTCTACCT ATATGCGTCA GTTAGCCATG ACGGCGGTTA TGGCCCAGCT	2160
GGGTTCCAT GTTCCTGCTG AAAGCGCCCA TTTACCGATT TTTGATGCGA TTTTACCCG	2220
TATCGGAGCA GCAGATGACT TGGTTTCGGG TCAGTCAACC TTTATGGTGG AGATGATGGA	2280
GGCCAATAAT GCCATTTCGC ATGCGACCAA GAACTCTCTC ATTCTCTTTG ATGAATTGGG	2340
ACGTGGAAC GCAACTTATG ACGGGATGGC TCTTGCTCAG TCCATCATCG AATATATCCA	2400
TGAGCACATC GGAGCTAAGA CCCTCTTTGC GACCCACTAC CATGAGTTGA CTAGTCTGGA	2460
GTCTAGTTTA CAACACTTGG TCAATGTCCA CGTGGCAACT TTGGAGCAGG ATGGGCAGGT	2520
CACCTTCCTT CACAAGATTG AACCGGGACC AGCTGATAAA TCTACGGTAT CCATGTTGCC	2580
AAGATTGCTG GCTTGCCAGC AGACCTTTTA GCAAGGGCGG ATAAGATTTT GACTCAGCTA	2640
GAGAATCAAG GAACAGAGAG TCCTCTCCC ATGAGACAAA CTAGTGCTGT CACTGAACAG	2700
ATTTCACCTT TTGATAGGGC AGAAGAGCAT CCTATCCTAG CAGAATTAGC TAAACTGGAT	2760
GTGTATAATA TGACACCTAT GCAGGTTATG AATGTCTTAG TAGAGTTAAA ACAGAACTA	2820
TAAAACCAAG ACTCACTAGT TAATCTAGCT GTATCAAGGA GACTTCTTTG ACAATTCTCC	2880
ACTTTTTTGC TAGAATAACA TCACACAAAC AGAATGAAAA GGAGCTGACG CATGTGCGCT	2940
CCCTTTTGTG TATTTTTTAA GGAGAAAGTA TGCTGATTCA GAAAATAAAA ACCTACAAGT	3000
GGCAGGCCCT GGCTTCGCTC CTGATGACAG GCTTGATGGT TGCTAGTTCA CTTCTGCAAC	3060
CGCGTTATCT GCAGGAAGTC TTAGGCGCCC TCCTTACTGG GAAATATGAA GCTATTTATA	3120
GTATCGGGGC TTGGTTGATT GGTGTGGCCG TAGTCGGTCT AGTTGCTGGT GGACTCAATG	3180
TTGTCTCTGC AGCCTATATT GCCCAAGGAG TTTTCATCCGA CCTTCGGGAG GATGCCCTCC	3240
GTAAAATTCA AACCTTTTCT TATGCTGATA TTGAACAATT TAATGCGGGA AATCTAGTCG	3300
TTCGAATGAC AAATGATATC AACCAGATTC AGAACGTTGT CATGATGACC TTCCAAATTC	3360
TTTTCAGACT TCCCCTCTTG TTCATCGGTT CGTTTATCCT AGCGGTTCAA ACCTTACCTT	3420
CTCTGTGGTG GGTGATTGTT CTCATGGTAG TCTTGATTTT TGGTTTGACT GCTGTCATGA	3480
TGGGAATGAT GGGCCTCGT TTGCCAAGT TTCAAACCTT TCTTGAGCGC ATCAATGCCA	3540
TTGCCAAGGA AAATTTACGT GGCCTTCGTG TGGTCAAGTC CTTTGTCCAA GAAAAAGAGC	3600
AATTTGCTAA GTTTACAGAG GTCTCAGACG AGCTTCTTGG TCAAAACCTT TACATTGGTT	3660
ATGCCCTTTC AGTAGTGAA CCCTTTATGA TGTTGGTTGG TTACGGGGCG GTCTTCTCTT	3720
CTATTTGGCT GGTGCGGGA ATGGTTCAGT CGGATCCGTC TGTTGTTGGT TCCATCGCTT	3780
CTTTTGTTAA TTACCTAAGC CAGATTATCT TTACCATTGT TATGGTTGGA TTTTGGGAA	3840
ATTCTGTCAG CCGTGCCATG ATTTCCATGC GTCGTATTCT AGAAATTCTT GACGCAGAGC	3900

535

CAGCTATGAC CTTCAAGGAT ATCCCAGATG AAGAGTTGGT TGGAAGTCTT AGCTTTGAAA	3960
ATGTGACCTT TACCTATCCA ATGGACAAGG AACCGATGCT GAAAGATGTG AGCTTTACTA	4020
TTGAACCTGG TCAAATGGTT GGTGTAGTTG GAGCGACTGG TGCAGGAAAG TCAACCTTGG	4080
CTCAATTGAT TCCACGTCTC TTTGATCCAC AGGACGGGGC CATTAAAAATC GGTGGCAAGG	4140
ATATTCGAGA AGTGAGTGAA GGAACCCTGC GTAAAACAGT TTCCATCGTT CTCCAACGTG	4200
CCATTCTTTT TAGTGAACG ATTCGAGATA ACTTGAGACA GGGGAAGGGG AATGCTACTC	4260
TATTTGAAAT GGAGCGCGCA GCCAATATTG CCCAGGCTAG TGAATTCATT CATCGTATGG	4320
AGAAAACCTT TGAAAGTCCA GTTGAAGAAC GGGGAACCAA TTTCTCTGGT GGACAAAAAC	4380
AAAGGATGTC GATTGCGCGT GGGATTGTCA GCAATCCACG TATTCTGATT TTGATGATT	4440
CGACCTCAGC CTTGGATGCC AAATCAGAGC GCTTGGTGCA AGAAGCTTTG AATAAGGACT	4500
TGAAGGGGAC GACAACCATT ATTATTGCTC AAAAAATTAG CTCGGTTGTC CATGCAGACA	4560
AGATCTTGGT TCTAAATCAA GGACGATTGA TTGGTCAAGG TACGCATGCA GACTTGGTTG	4620
CCAACAATGC CGTTTACCGT GAAATCTATG AAACACAGAA ATGAAAGACA AACTATAAGA	4680
AAAGTCAATA GTTTTATCTA AACTATTTCT TATTTCAATT TGATGATTTG GCGATGATTT	4740
TAGAGCACGG CAAAAAGCCC TTGAAAAAGT CCATTTTTC AAAGGTAATC CTGTGTTAAT	4800
TTCAGAAAT ACATCACTTT TTGTTGCTCA AATGGCAGCT CTTTTTTTAG GATATAAAAC	4860
AGGGTTCGGA TAAGTTTTTT TGCAAGGTGG ATGATGGCTA CATTGTAATG TTTTCCTTGT	4920
TCTAATTTAG TCTTAAGATA GGCCTTAAAA GCAGGCGAAA AGCGAGGGCA TGCTTTGGCA	4980
GCTTGATGA GTACCTACCG CAGATGAGGG GAACTCCGTT TGACCATTCT TCCTGCTAAA	5040
TCAATCTGAT CTGACTGATA AATAGAAGAA TCCAGTCCAG CGAAAGCTTG TAATTGAGCA	5100
GGATTATCAA AGGCATGAAT ATTTGGAATC TCAGCTAAAA TGACCGCCCC TAAACGATCC	5160
CCAATCCCAG TAACCGTCGT GATGACCGAG TTGAACTCAG CCATCAAGTC ATTGACACAT	5220
GTTCGCGCCT TGTCAATGAG CCTCTGTAA TGTGTGATGT TTTCATTACA CGAGATAAAA	5280
CGTCTATGCG TTATCAAACCT CATTACCAAT TAAAACAAAA AGCTGTGGTT AGATCCTTTC	5340
GGAAATTGTC AAGCGATTGG AGGAAATGAA CTAATCCACA GCGGCTTATT CCAAGTATAC	5400
CACTTGGGCT TTGGCAGTAG CTAAGTCCGC TAAATATAAT ATAAGGAGGA GTAAAAAGAA	5460
GACAGTTCAA TTTTTTTGGC ATTATTTTAA GGTCTACAAG TTCTCATTTG TAGTTGTCAT	5520
CCTGATGATT GTTCTGGCGA CTTTGTGCGA AGCCCTCTTT CCAGTCTTTT CTGGACAAGC	5580
GGTGACGCAG CTAGCCAATT TAGTTCAAGC TTATCAAAAT GGCAATCCAG AACTTGTATG	5640

536

GCAAAGCCTA TCAGGAATCA TGGTCAATCT TGGCCTGCTG GTTTTGGTTC TATTTATCTC	5700
TAGTGTAATA TACATGTGTC TCATGACGCG CGTGATTGCA GAATCGACCA ACGAGATGCG	5760
CAAAGGCCTC TTTGGTAAGC TTGCTCAGTT GACGGTTTCT TTCTTTGACC GTCGACAAGA	5820
TGGCGATATC CTGTCTCATT TTACCAGTGA TTTGGATAAT ATCCTCCAAG CCTTTAACGA	5880
AAGCTTGATT CAGGTCATGA GCAATATTGT TTTATACATT GGTCTGATTC TTGTCATGTT	5940
TTGAGAAAT GTGACGCTGG CTCTCATCAC CATTGCCAGC ACCCCATTGG CTTTCCTTAT	6000
GCTGATTTTC ATCGTGAAAA TGGCAGCAA ATACACCAAC CTCCAGCAGA AAGAGGTAGG	6060
GAAGCTCAAC GCCTATATGG ATGAGAGCAT CTCAGGCCAA AAAGCCGTGA TTGTGCAAGG	6120
AATCAAGAG GATATGATGG CAGGATTTCT TGAACAAAAT GAGCGCGTGC GCAAGGCAAC	6180
CTTTAAAGGA AGAATGTTCT CAGGAATTCT TTTCCCTGTC ATGAATGGGA TGAGCCTGAT	6240
TAATACAGCC ATCGTCATCT TTGCTGGTTC GGCTGTACTT TTGAATGATA AGTCTATTGA	6300
AACAAGTACA GCCCTAGGTT TGATTGTTAT GTTTGCACAA TTTTCACAGC AGTACTACCA	6360
GCCTATTATC CAAGTTGCAG CGAGTTGGGG AAGCCTTCAG TTGGCCTTTA CTGGAGCTGA	6420
ACGAATTCAG GAAATGTTTG ATGCAGAGGA GGAAATCCGA CCTGAAAAGG CTCCAACCTT	6480
CACTAAGTTG CAAGAAAGTG TTGAAATCAG TCATATCGTT TTTTCATACT TGCCTGATAA	6540
ACCTATTTTG AAAGATGTCA GCATTTCTGC CCCTAAAGGC CAGATGACAG CAGTTGTTGG	6600
GCCGACAGGT TCAGGAAAAA CGACTATTAT GAACCTCATC AATCGCTTTT ATGATGTTGA	6660
TGCTGGTGGT ATTTATTTTG ATGGTAAAGA CATTCTGGC TATGACTTAG ATAGTCTTAG	6720
AAGCAAGGTG GGAATTGTAT TGCAAGATTC GGTCTTGTTT AGCGGAACGA TTAGAGACAA	6780
TATCCGATTT GGTGTGCCAG ATGCTAGTCA GGAAATGGTT GAGGTAGCAG CAAAAGCAAC	6840
CCACATTCAC GACTATATCG AAAGTTTGCC TGATAAGTAC GATACTCTTA TTGATGATGA	6900
CCAGAGCATC TTTTCAACAG GGCAGAAGCA ATTGATTTC ATCGCTCGAA CCCTGATGAC	6960
AGATCCAGAA GTTCTCATTC TCGATGAAGC AACTTCAAAC GTAGATACGG TGACAGAAAG	7020
CAAGATTCAG CATGCCATGG AGGTGGTTGT AGCAGGTAGA ACTAGTTTCG TCATTGCCCA	7080
CCGCTTGAAA ACCATTCTCA ATGCAGATCA GATTATTGTC CTTAAAGATG GAGAAGTCAT	7140
TGAACGTGGT AACCACCATG AACTTTTGAA GCTAGGTGGC TTTTATTCAG AACTCTATCA	7200
CAATCAATTT GTTTTCGAAT AAGAAAGAAG TTGTCTATG TGGGCAGCTT TTTCTGTCC	7260
ATAAAAAATG TTTATCACAG CCTTAAAAA AACATATTAG ACGAAAGTCA TTTTGAGTGA	7320
TATGATAGGA CTATCGTTAG CATTCGAAAG GAGAGGCATC ATGGCTAGAA CGGTGTAGG	7380
AGTTGCTGCA AATCTATGTC CCGTAGACGC AGAAGGCAAA ATCATTCATT CATCTGTATC	7440

537

TTGTAGATTC GCAGAGATCA TTCGTCAAGT CGGTGGTCTC CCTTTAGTCA TTCCTGTTGG	7500
TGATGAGTCA GPTGTACGTG ATTATGTGGA AATGATTGAC AAACATCATTT TGACAGGAGG	7560
CCAAAATGTT CATCCTCAGT TTTATGGAGA GAAAAAGACC GTCGAGAGCG ATGATTACAA	7620
TCTGGTCCGT GACGAATTTG AATGGGCACT CTTGAAGGAA GCGCTTCGTC AGAATAAACC	7680
AATTATGGCA ATCTGTGCGG GTGTCCAAC TGTCAATGTT GCCTTTGGTG GAACCCCTCAA	7740
TCAAGAAATC GAAGGTCAGG	7760

(2) INFORMATION FOR SEQ ID NO: 64:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2723 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

GAGGTTTTAA TTCACCTACC TCTsCCGTAT CTTTATTTAA AATGAATTCT TTTACGGTTG	60
TATTTCTTGC AAAATCTTTT ACAACAATCT TAATGTTTAG TGTCTTGTCT ATTATTTGTT	120
TAATATCATT AAATGATGTA TATCTTTTC CATTATATA AATATGTTGT TCTTGAATCT	180
CACCATCGAA TCCATTATTT CTTTATCAT TGATGTTAAA GACTACAGAT TTTCCATCAG	240
CATATTCGAT ACTAGTATTT CCCTTAGGAT CAATGTTTAC TTCGGGTTTA ACATTATCAT	300
ATAAAAACTG ATAGTGGACT CCAACTGCTT TAGCATTCAA ATCGCTATAG CCAGTTTGAA	360
GATAAACATT TCCATCCATA TCTGTTACCT TATCTGGAAA TCCGTTTGCT TTATAGTCTT	420
TCATTCCCCA GTCCATGATG TCACCGTCTT TAACATTCAG CTTAATATTA AAATCTCTAG	480
TGTTATCAAT GTGTAAATCT CCGTAGATTA AATAATTATC TACAACCGAT TCATTAATC	540
TCAATCCCCA GTTAAAACCA CCCTTATCAG AAATCTTACC TCTTAAATAA AATTCTGGAT	600
TTCGTACATA AATTTTATTA GATTTAGATG GATTAAAGTA GTTCTTATCC ATTGAAAGGT	660
TTACTGGTTT GGTATCAATA AATAACATGG AGCCATCTTC TTTTATAGCT TCTACATTGA	720
ACTTATCCTC TCCAGTGTAT TCTTTATCAT CCTTACCAA TAATACAAGT TTAGAAGAAT	780
CTGTCACAAG ATTTCCGTCT TTATCGATAG CTTCCCCCTT ATCGTTCATT TTAAATGTAA	840
ACACTTGATA CCTTATAATG TTAAAGCCGT CCAAAGCCGA CATTAAATACA GATTGGGTAC	900
TTCTTCCATC TTCAACATTT CTAATATCAG CATAAATTGT TGTTTCTGAA AGGGCTCTTA	960
GATTAGGATT GGCCTTTTGT ATTTTGTGCTA TATCTTCCTT GCTATAGACT CCATTTCTCT	1020

538						
CTAACATATC	CGTTTTTCCA	GGATTATAGG	TAGTCACTTT	TAGTGCATAG	CCTTTTCTTA	1080
GAATGATATT	ATCCTTTAAC	AGATATTGTT	GTTTTTCTGA	ATCAGAATAG	ATTTTACCAG	1140
ATTCCATTTT	AGTTAAATTG	TCTGGTTTGT	TTTTTGAAAG	ATCTCCTTCC	CCTAATTCTTA	1200
TGACATTCCC	ATAACTTGAT	ACATAGGGAT	ATTCTGATTT	AGTTTCTCTTA	ATTTTTTTCAG	1260
GCATTCTAAT	TPTAATTTCA	GCTTTTTTCT	GATCATTATC	TTTAACAAAT	AATCTCATAT	1320
CTCCTGCAAA	AGCTAATCCA	TCCACAATAT	CATTAATATT	AGCGTATAGA	TCAAATGTCA	1380
TCGTTTTTGA	GTGGAAATCA	TACTTGGTCG	CTTTGATTTT	TATAGATTTA	TAGTTATTCC	1440
CATAATATAC	CTTGGCATTT	TTAGAAACAT	TACTTATCTT	TCCAAGAATT	TCAAAGTGTC	1500
CATCTTTAGA	CGGACTTAGA	ACACCATAAA	TTTTTGATTT	GATTTTCGTCA	AGTTTCTCAG	1560
TTTCATATTC	TAGATCAGTC	CCATCATCGT	AGGCTATTAT	ATTTCTCTTA	TCATCGTATT	1620
TATAATCGTA	TTCCTCCATT	CTCTTACCAG	TTTCACCTGT	AAAATCATCA	ACTTCTCTAA	1680
ATTTCTTTTT	AATGAGTTTC	TTTAAGTCTT	TATTTTCAAA	GTCTCTAATT	GTGAAATAT	1740
TTCTATCAAT	AGTAAACTA	GATTTTTTCTT	TAATAGACTC	TTCATTTTCT	TGATGATGAT	1800
GTTCTACCCC	AGTTGTATCT	TTTTTTAGAC	TACCCTCTTT	TCCATTTCTT	AAATTTTTAA	1860
ATTTAGATTC	TGCAATCTCG	CCAAGCTTTT	GATATTTAGA	TGAATCTTGA	TCAGGATCTA	1920
CTAGATAATA	GGAAATCATC	CCCTTTTCAT	CAGCCTGATT	AGCAAATTTA	ATTCTATGAA	1980
TCTTTGTGAA	ATTGCTAGAA	CCATCTAATG	CAATGACTTC	AATGATTTTT	CCCCTTAAAT	2040
CTCCCGCACC	TTTAATTTCA	TAAATGGTAT	TTCCGTCTTT	ATCAAGTTTT	CTATTTCTTC	2100
CTTGACCCCTC	ACCTGCGTAA	GTTACTTCAA	GATTTTTTTC	AACCTCTCCA	TCTTCATTAA	2160
CAAGAGCGGC	GCCAGCATAC	CAAACCTTCGT	TCGCAATCTC	GTCAAATTTT	TCAGGATGTT	2220
CTTTTGTGATC	TCTCGCAAAT	AGCGTTTCAT	TCTTATACTG	ATCTTTTACC	TTATGATAAG	2280
TATCCTTTGT	AATCAACTTA	ATTTTTTCAG	GATTTGAAAA	ATCAACCGAA	ACAATCTTAG	2340
GGGCGGTGTT	ATCAATTTTT	ACAGGAATAT	AGGAAACCTG	CCATGGGTAA	TCTTTAGTTA	2400
ATCTATATTT	AAATTTATAG	AAATATTGAC	CTTCCGCAAT	CGGTTCAAAT	TGACCTCTTA	2460
TCTTAGTAGC	AGGATCTTGA	TTATCCTTAC	TTTCTGGTGC	ATTTTCTTCT	CTACCTCTAG	2520
GATTATAGAT	GAGTCCATCC	CACCTCAAGT	CACCCCAAAC	TTTTAGTTTA	GATGATTTGA	2580
TTCCCTTTGC	ATCATTGCTT	TTAGAATTTA	AAATTCCTCT	AATAAAGTGT	TCTCTCGAAA	2640
TGACTTTTAA	GTCTCTTTGA	TTTTCTCCCT	CTTTATTTGT	ATTTACTATT	GAAATCAATC	2700
CTTCTTCTGC	ACTTCTTAAT	ACA				2723

(2) INFORMATION FOR SEQ ID NO: 65:

539

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11831 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

AAAAAAGTGG GAATGACTCA AATCTTCACT GAAGCTGGCG AATTGATCCC TGTAACAGTT	60
ATTGAAGCAA CTCCAAACGT TGTTCCTCAA GTTAAACTG TTGAAACAGA CGGATACAAC	120
GCTATCCAAG TTGGTTTCGA TGACAAACGC GAAGTATTGA GCAACAAACC TGCTAAAGGA	180
CATGTAGCGA AAGCTAACAC GGCTCCTAAG CGCTTCATTC GTGAATTCAA AAACGTTGAA	240
GGCTTGGAAG TTGGTGCTGA AATTACAGTT GAAACATTCG CAGCTGGAGA CGTTGTTGAC	300
GTAACGGGTA CTTCTAAAGG TAAAGGTTTC CAAGGTGTTA TCAAACGCCA CGGACAATCA	360
CGTGGACCAA TGGCTCACGG TTCTCGTTAC CACCGTCGTC CAGGTTCTAT GGGGCCTGTT	420
GCACCTAACC GCGTATTCAA AGGTAAAAAC CTTGCAGGAC GTATGGGTGG CGACCGCGTA	480
ACAATTCAAA ACCTTGAAGT TGTACAAGTT GTTCCAGAAA AGAACGTTAT CCTTATCAAA	540
GGTAACGTAC CAGGTGCTAA GAAATCTCTT ATCACTATCA AATCAGCAGT TAAAGCTGGT	600
AAATAATAAA GAAAGGGGAA ATCAGTCACA ATGGCAAACG TAACATTATT TGACCAAAC	660
GGTAAAGAAG CTGGCCAAGT TGTTCCTAGC GATGCAGTAT TTGGTATCGA ACCAAATGAA	720
TCAGTTGTGT TTGATGTAAT CATCAGCCAA CGCGCAAGCC TTCGTCAAGG AACACACGCT	780
GTTAAAAACC GCTCTGCAGT ATCAGGTGGT GGACGCAAAC CATGGCGTCA AAAAGGAACT	840
GGACGTGCTC GTCAGGTTTC TATCCGCTCA CCACAATGGC GTGGTGGTGG TGTGTCTTC	900
GGACCAACTC CACGTTTATA CGGCTACAAA CTTCCACAAA AAGTTCGTCG CCTAGCTCTT	960
AAATCAGTTT ACTCTGAAAA AGTTGCTGAA AACAAATTCG TAGCTGTAGA CGCTCTTTCA	1020
TTTACAGCTC CAAAAACTGC TGAATTGCA AAAGTTCTTG CAGCATGAG CATCGATTCT	1080
AAAGTTCTTG TTATCCTGA AGAAGGAAAT GAATTCGCAG CTCTTTCAGC TCGTAACCTT	1140
CCAAACGTGA AAGTTGCAAC TGCTACAAC GCAAGTGTTT TTGACATCGC AAATAGCGAC	1200
AAACTTCTTG TCACACAAGC AGCTATCTCT AAAATCGAGG AGGTTCTTGC ATAATGAATT	1260
TGTATGATGT TATCAAAAAA CCTGTCACTA CTGAAAGCTC AATGGCTCAA CTTGAAGCAG	1320
GAAATATGT ATTTGAAGTT GACACTCGTG CACACAACT TTTGATCAAG CAAGCTGTTG	1380
AAGCTGCTTT CGAAGGTGTT AAAGTTGCCA ATGTTAACAC AATCAACGTA AAACCAAAAG	1440

540

CTAAACGTGT TGGACGTAC ACTGGTTTCTA CTAACAAAAC TAAAAAAGCT ATCATCACAC	1500
TTACAGCTGA TTCTAAAGCA ATCGAGTTGT TTGCTGCTGA AGCTGAATAA TCTAAGGAGG	1560
AAATATCGTG GGAATTCGTG TTTATAAACC AACACAAAAC GGTCGCCGTA ATATGACTTC	1620
TTTGGAATTC GCTGAAATCA CAACAAGCAC TCCTGAAAAA TCATTGCTTG TTGCATTGAA	1680
GAGCAAGGCT GGTCGTAACA ACAACGGTCG TATCACAGTT CGTCACCAAG GTGGTGGACA	1740
CAAACGTTTC TACCGTTTGG TTGACTTCAA ACGTAATAAA GACAACGTTG AAGCAGTTGT	1800
TAAAACAATC GAGTACGATC CAAACCGTTC TGCAAAACATC GCTCTTGCTAC ACTACACTGA	1860
CGGTGTGAAA GCATACATCA TCGCTCCAAA AGGTCTTGAA GTAGGTCAAC GTATCGTTTC	1920
AGGTCCAGAA GCAGATATCA AAGTCGGAAG CGCTCTTCCA CTTGCTAACA TCCCAGTTGG	1980
TACTTTGATT CACAACATCG AGTTGAAACC AGGTCTGGT GGTGAATTGG TACGTGCTGC	2040
TGGTGCATCT GCTCAAGTAT TGGGTCTGA AGGTAAATAT GTTCTTGTTT GTCTTCAATC	2100
AGGTGAAGTT CGTATGATTC TTGGAAGTTG CCGTGCTACA GTTGGTGTG TCGGAAACGA	2160
ACAACATGGA CTTGTAAACC TTGGTAAAGC AGGACGTAGC CGTTGGAAAG GTATCCGCCC	2220
AACAGTTCTG GGTCTGTAA TGAACCTAA CGATCACCCA CACGGTGGTG GTGAAGGTAA	2280
AGCACCAAGT GGTCGTAAAG CACCATCTAC TCCATGGGGC AAACCTGCTC TTGGTCTTAA	2340
AACTCGTAAC AAGAAAGCGA AATCTGACAA ACTTATCGTT CGTCGTCGCA ACGAGAAATA	2400
ATATTAAACT AGTCGCTTAA GCAACTAGTA AATCCGCCAG CTCGGTAGCG CTCCATAGGA	2460
GTGCAAGCCG CTGTGGTACA ACATTTAAAG GAGAAAATAT AAAAATGGGA CGCAGCTTAA	2520
AAAAAGGACC TTTCGTCGAT GAGCATTGTA TGAAAAAAGT TGAAGCTCAA GCTAACGACG	2580
AAAAGAAAAA AGTTATTAAA ACTTGGTCAC GTCGTTCAAC GATCTTCCCA AGTTTCATTG	2640
GTTACACTAT TGCAGTTTAT GACGGACGTA AACACGTACC TGTTTACATC CAAGAAGACA	2700
TGGTAGGCCA CAAACTTGGT GAATTTGCAC CAACTCGTAC TTACAAAGGT CACGCTGCAG	2760
ACGACAAGAA AACACGTAGA AAATAAGGAG AACATAAATG GCAGAAATTA CTTAGCTAA	2820
AGCAATGGCT CGTACAGTAC GTGTTTCACC TCGTAAATCA CGTCTTGTTT TTGATAACAT	2880
CCGTGGTAAA AGCGTAGCCG ATGCAATCGC AATCTTGACA TTCACTCCAA ACAAAGCTGC	2940
TGAAATCATC TTGAAAGTTT TGAAGTCAGC TGTAAGCTAAC GCTGAAAACA ACTTTGGTTT	3000
GGATAAAGCT AACTTGGTAG TATCTGAAGC ATTCGCAAAC GAAGGACCAA CTATGAAACG	3060
TTTCCGTCCA CGTGCAGAAAG GTTCAGCTTC ACCAATCAAC AAACGTACAG CTCACATCAC	3120
TGTAGCTGTT GCAGAAAAAT AAGGAGGTAA AATCGTGGGT CAAAAAGTAC ATCCAATTGG	3180
TATGCGTGTC GGCATCATCC GTGATTGGGA TGCCAAATGG TATGCTGAAA AAGAATACGC	3240

541

GGATTACCTT CATGAAGATC TTGCAATCCG TAAATTCGTT CAAAAGAAGC TTGCTGACGC	3300
AGCAGTTTCA ACTATTGAAA TCGAACGCGC AGTAAACAAA GTTAACGTTT CACTTCACAC	3360
TGCTAAACCA GGTATGGTTA TCGGTAAAGG TGGTGCTAAC GTTGATGCAC TCCGTGCAAA	3420
ACTTAACAAA TTGACTGGAA AACAACTACA CATCAACATC ATCGAAATCA AACAACTGA	3480
TTTGATGCT CACCTTGTAG GTGAAGGAAT TGCTCGTCAA TTGGAGCAAC GTGTTGCTTT	3540
CCGTCTGCA CAAAACAAG CAATCCAACG TGCAATGCGT GCTGGAGCTA AAGGAATCAA	3600
AACTCAAGTA TCAGGTCGTT TGAACGGTGC AGATATCGCC CGTGCTGAAG GATACTCTGA	3660
AGGAAGTGT CCGCTTCACA CACTTCGTGC AGATATCGAT TACGCTTGGG AAGAAGCAGA	3720
TACTACATAC GGTAAGCTTG GTGTTAAAGT ATGGATCTAC CGTGGTGAAG TTCTTCCAGC	3780
TCGTAAAAAC ACTAAGGAG GTAAATAACC AATGTTAGTA CCTAAACGTG TTAAACACCG	3840
TCGTGAGTTC CGTGAAAAA TGCGCGGTGA AGCAAAAGGT GGAAAAGAAG TAGCATTCGG	3900
TGAATACGGT CTTCAGCTA CAACTAGCCA CTGGATCACT AACCGCCAAA TCGAAGCTGC	3960
TCGTATCGCC ATGACTCGTT ACATGAAACG TGGTGGTAAA GTTTGGATTA AAATCTTCCC	4020
ACACAAATCA TACACTGCTA AAGCTATCGG TGTGCGTATG GGATCTGCTA AAGGGGCACC	4080
TGAAGGTTGG GTAGCACCAG TTAAACGTGG TAAAGTGATG TTCGAAATCG CTGGTGTATC	4140
TGAAGAGATT GCACGTGAAG CGCTTCGACT TGCTAGCCAC AAATTGCCAG TTAAATGTAA	4200
ATTCGTAAAA CGTGAAGCAG AATAAGGAGA AGGCATGAAA CTAAATGAAG TAAAGAATT	4260
TGTTAAAGAA CTTCTGGTC TTTCTCAAGA AGAACTCGCG AAGCGCGAAA ACGAATTGAA	4320
AAAAGAATTG TTTGAACTTC GTTCCAAGC TGCTACTGGT CAATTGGAAC AACAGCTCG	4380
CTTGAAAGAA GTTAAAAAAC AAATCGCTCG CATCAAAACA GTTCAATCTG AAGCGAAATA	4440
ATAGACTAGG GAAGGAGAAA TTTCATGGA ACGCAATAAT CGTAAAGTTC TTGTTGACG	4500
TGTTGTATCT GACAAAATGG ACAAGACAAT CACAGTTGTA GTTGAAACAA AACGTAACCA	4560
CCCAGTCTAT GGTAAACGTA TTAAGTACTC TAAAAAATAC AAAGCTCATG ATGAAAACAA	4620
TGTTGCCAAA GAAGGCGATA TCGTACGTAT CATGGAAACT CGCCCGCTTT CAGCTACAAA	4680
ACGTTTCCGT CTTGTAGAAG TTGTTGAAGA AGCGGTCATC ATCTAATCAA ACCTGAAAGG	4740
AGAAAACTGA AATGATTCAA ACAGAACTC GTTTGAAAGT CGCAGACAAC AGCGGTGCTC	4800
GCGAAATCTT GACTATCAAA GTTCTTGGTG GTTCAGGACG TAAATTGCA AACATCGGTG	4860
ATGTTATCGT GGCATCTGTA AAACAAGCTA CTCCTGGTGG TCGGTTAAA AAAGGTGACG	4920
TTGTTAAAGC AGTTATCGTT CGTACTAAAT CAGGTGCTCG TCGTGCTGAT GGTTCATACA	4980

542

TCAAATTTGA CGAAAACGCA GCAGTTATCA TCCGTGAAGA CAAAACCTCCT CGCGGAACAC 5040

GTATCTTTGG CCCAGTTGCA CGTGAATTGC GTGAAGGTGG CTTTCATGAAG ATCGTGTAC 5100

TTGCTCCAGA AGTACTTTAA TTTTAGGAA CAACTAGTC CCCTAGCTTC AAGCTAGGGT 5160

CCCCTTATGG GCGTAAGAAA AATCAAGGAG AAACCTAATG TTTGTAAAAA AAGGCGACAA 5220

AGTTCGCGTA ATCGCTGGTA AAGATAAGGG AACAGAAGCT GTGTCTCTTA CTGCCCTTCC 5280

AAAAGTAAAC AAAGTTATCG TTGAAGGTGT TAACATTGTT AAGAAACACC AACGTCCAAC 5340

TAACGAGCTT CCTCAAGGTG GTATCATCGA GAAAGAAGCA GCTATCCACG TATCAAACGT 5400

TCAAGTTTGG GACAAAAATG GTGTAGCTGG TCGTGTGGA TACAAATTTG TAGACGGTAA 5460

AAAAGTTCGC TACAACAAAA AATCAGGCGA AGTGCTTGAT TAATCACGAA GGAAAGGAGA 5520

AGTATAATGG CAAATCGTTT AAAAGAAAAA TATCTTAATG AAGTAGTCC TGCTTTGACA 5580

GAACAATTCA ACTACTCATC AGTGATGGCT GTGCCTAAAG TAGATAAGAT TGTTTTGAAC 5640

ATGGGTGTTG GTGAAGCTGT ATCAAACGCT AAAAGCCTTG AAAAAGCTGC TGAAGAATTG 5700

GCACTTATCT CAGGTCAAAA ACCACTTATC ACTAAAGCTA AAAATCAAT CGCCGGCTTC 5760

CGTCTTCGTG AAGGTGTTGC GATCGGTGCA AAAGTTACCC TTCGTGGTGA ACGTATGTAC 5820

GAATTCCTGG ATAAATTGGT ATCAGTTTCA CTCCACGTG TACGTGACTT CCACGGTGTG 5880

CCAACAAAAT CATTGATGG ACGCGGGAAC TACACACTTG GTGTGAAAGA ACAATTAATC 5940

TTCCCAGAAA TCAACTTCGA TGACGTGAC AAAACTCGTG GTCTTGACAT CGTTATCGTA 6000

ACAACTGCTA AACTGACGA AGAGTCACGT GCATTGCTTA CAGGCCTTGG AATGCCTTTT 6060

GCAAAATAAT ATAGGAGTA AATCTAATGG CTAAAAATC AATGGTAGCT AGAGAGGCTA 6120

AACGCCAAAA AATTGTTGAC CGTTATGCTG AAAACGTGC TGCATTAAAG GCGGCAGGGG 6180

ACTACGAAGG TTTATCTAAA TTACCTCGCA ACGCCTCACC GACTCGTTTA CATAATCGTT 6240

GTAGGGTTAC GGGCGCCCA CATTCACTTT ACCGCAAATT TGGTCTGAGT CGTATCGCTT 6300

TTCCGGAAC TCGCATAAA GGTCAAATTC CTGGTGTAAC AAAAGCATCT TGGTAATTTA 6360

AGATATCAAG AGCGTCAAAA CTCCAAGTAA AAATAGGAAA CTGACGAAG AACTAAAGT 6420

TTCTAGGAAA GTTTATCTTT TTCACACAGA GTTTAGCCCG GGTTCATTG GGCTTGCCAA 6480

TTTGAACACG AGCTACAGCT TTGGCAAAAA AGACCAATTT GCTTTGGAGC ATTGCTTCTG 6540

CATTAAATTG TCTATTTTGG CTCGTGCTGT TACGCTCTTT GTATCATGTA TTAAC TAGCA 6600

AGTGCAACTT GCAAACTACT AGTAAGAGGA GAAAAACAAA ATGGTTATGA CTGACCCAAT 6660

CGCAGACTTC CTAACGTA TTCGTAATGC TAACCAAGCT AAACACGAAG TACTTGAAGT 6720

ACCTGCATCA AACATCAAAA AAGGATTGC TGAAATCCTT AAACGCGAAG GTTTTGTA 6780

543

AAACGTTGAA ATCATTGAAG ATGACAAACA AGGCGTCATC CGTGTATTTC TTAAATACGG	6840
ACCAAATGGT GAGAAAGTTA TCACTAACTT GAAACGTGTT TCTAAACCAG GACTTCGTGT	6900
CTACAAAAAA CGTGAAGACC TTCCAAAAGT TCTTAACGGA CTTGGAATTG CCATCCTTTC	6960
AACTTCTGAA GGTTCGCTTA CTGATAAGA AGCACGCCAA AAGAATGTTG GTGGTGAGGT	7020
TATCGCTTAC GTTTGGTAAA ATCAAGATAC AAAGCTCGTA AAGAACAAAG CAAAATTAGG	7080
AAGTTGGAGA AGTTTGTTTA CAAACAAGCC AACTTATCTA TTTTGCACAG TTCTTAGAGC	7140
GTGTTTCAGTT CAGCTCTTGA ACTAAATAAG TATCTGAACC CCGTGAAAAC TGGCCGTTCT	7200
GGCCTGACAA TTTAACAGGA GAAATAAAC ATGTACCGTA TTGGTAATAA AGTTATCGTG	7260
TTGCCTGCTG GTGTTGAAC CTCTAACAAT GACAACGTTG TAACTGTAAA AGGATCTAAA	7320
GGAGAAGTTA CTCGTGAGTT CTCAAAGAT ATTGAAATCC GTGTGGAAGG TACTGAAATA	7380
ACTCTTCACC GTCCAAACGA TTCAAAGAA ATGAAAACTA TCCACGGAAC TACTCGTGCC	7440
CTTTTGAACA ACATGGTTGT TGGTGTATCA GAAGGATTCA AGAAAGAACT TGAATGCGT	7500
GGGGTTGGTT ACCGTGCACA GCTTCAAGGA TCTAAACTTG TTTTGGCTGT TGTAAATCT	7560
CATCCAGACG AAGTTGAAGC TCCAGAAGGA ATTACTTTTG AACTTCCAAA CCCAACAACA	7620
ATCGTTGTTA GCGGAATTTT AAAAGAAGTA GTTGTCTAAA CAGCTGCTTA CGTACGTAGC	7680
CTTCGTTTAC CAGAACCATA TAAAGGTAAA GGTATCCGTT ACGTTGGTGA ATTCGTTTCG	7740
CGTAAAGAAG GTAAACAGG TAAATAATGT TGAGTGGTTG ATCATCAACC ACCAACCTAT	7800
TTTCCAACTT TGTGCATAGC ACACGATTTA AAATAAAGA GGTGAAAACT GTGATTTCAA	7860
AACCAGATAA AAACAACTC CGCCAAAAAC GCCACCGTCG CGTTCGCGGA AAATCTCTG	7920
GAATGCTGA TCGCCACAGT TTGAACGTAT TCCGTTCTAA TACAGGCATC TACGCTCAAG	7980
TGATTGATGA CGTAGCGGT GTAACGCTCG CAAGTGCTTC AACTCTTGAT AAAGAAGTTT	8040
CAAAAGGAAC TAAACTGAA CAAGCCGTTG CTGTCGGTAA ACTCGTTGCA GAACGTGCAA	8100
ACGCTAAAGG TATTTAGAA GTGGTGTTTC ACCGCGGTGG ATATCTATAT CACGGACGTG	8160
TGAAAGCTTT GGCTGATGCA GCTCGTGAAA ACGGATTGAA ATTCTAATAG GAGGACACTA	8220
GAAAATGGCA TTAAAGACA ATGCAGTTGA ATTAGAAGAA CGCGTAGTTG CTGTCAACCG	8280
TGTTACAAAA GTTGTAAAG GTGGACGTCG TCTTCGTTTC GCAGCTCTTG TTGTTGTTGG	8340
TGACCACAAT GGTGCGGTAG GATTTGGTAC TGGTAAAGCT CAAGAAGTTC CAGAAGCAAT	8400
CCGTAAAGCA GTAGATGATG CTAAGAAAAA CTTGATCGAA GTTCCTATGG TTGGAACAAC	8460
AATCCACAC GAAGTTCTTT CAGAATTCGG TGGAGCTAAA GTATTGTTGA AACCTGCTGT	8520

544

AGAAGGTTCT GGAGTTGCCG CTGGTGGTGC AGTTCGTGCC GTTGTGGAAT TGGCAGGTGT	8580
GGCAGATATT ACATCTAAAT CACTTGGTTC TAACACTCCA ATCAACATTG TTCGTGCAAC	8640
TGTTGAAGGT TTGAAACAAT TGAAACGCGC TGAAGAAAT GCTGCCCTTC GTGGTATTTC	8700
AGTTTCTGAT TTGGCATAAG AAAGGGGATA AAATGGCTCA AATTAAAATT ACTTTGACTA	8760
AGTCTCCAAT CGGACGCATT CCATCACAAC GTAAACTGT TGTAGCACTT GGACTTGGCA	8820
AATTGAACAG CTCTGTTATT AAAGAAGATA ACGCTGCTAT CCGTGGTATG ATCACAGCAG	8880
TATCTCACTT AGTAACAGTT GAAGAAGTAA ACTAATGAAG TTTTAGGGGA TGTGCACTGT	8940
ACCATCCCCT AAAACTAGAT ATAGTCATCT ATGATGACAT CGTATAGGCG AGTTGATGGG	9000
GGAGACAACC TTTTCTCCCT TATCGGCGCT AGCATTTTAC AAAAGAGGAG AAAATAAAAA	9060
TGAAACTTCA TGAATTGAAA CCTGCAGAAG GTTCTCGTAA AGTACGTAAC CGCGTTGGTC	9120
GTGGTACTTC ATCAGGTAAC GGTAAAACAT CTGGTCGTGG TCAAAAAGGT CAAAAGCTC	9180
GTAGCGGTGG CGGAGTTCGC CTTGGTTTGG AAGGTGGACA AACTCCATTG TTCCGTCGTC	9240
TTCCAAAACG TGGATTCACT AACATCAACG CTAAAGAATA CGCAATTGTG AACCTTGACC	9300
AATTGAACGT CTTGAAGAT GGTGCTGAAG TAACTCCAGT TGTCTTATC GAAGCAGGAA	9360
TTGTTAAAGC TGAAAAGTCA GGTATTAAAA TTCTTGGTAA CGGTGAGTTG ACTAAGAAAT	9420
TGACTGTGAA AGCAGCTAAA TTCTCTAAAT CAGCTGAAGA AGCTATCACT GCTAAAGGTG	9480
GTTCAGTAGA AGTCATCTAA GAGAGGTGAC CTATGTTTTT TAAATTATTA AGAGAAGCTC	9540
TTAAAGTCAA GCAGGTCGA TCAAAAATTT TATTTACAAT TTTTATCGTT TGGTCTTTT	9600
GTATCGGAAC TAGCATTACA GTTCTGGTG TGAATGCCAA TAGCTTGAAT GCTTTAAGTG	9660
GATTATCCTT CTTAAACATG TTGAGCTTGG TGTGGGGAA TGCCCTAAAA AACTTTTCGA	9720
TTTTTGCCCT AGGAGTTAGT CCCTATATCA CCGCTTCTAT TGTGTCCAA CTCTTGCAA	9780
TGGATATTTT ACCCAAGTTT GTAGAGTGGG GTAAACAAGG GGAAGTAGGT CGAAGAAAAT	9840
TGAATCAAGC TACTCGTTAT ATTGCTCTAG TTCTCGCTTT TGTGCAATCT ATCGGGATTA	9900
CAGCTGGTTT TAATACCTTG GCTGGAGCTC AATTGATTAA AACTGCTTTA ACTCCACAAG	9960
TTTTTCTGAC GATTGGTATC ATCTTAACAG CTGGTAGTAT GATTGTCACT TGGTTGGGTG	10020
AGCAAATTAC AGATAAGGGA TACGGAAACG GTGTTTCCAT GATTATCTTT GCCGGGATTG	10080
TTTCCTCAAT TCCAGAGATG ATTCAGGGCA TCTATGTGGA CTACTTTGTG AACGTCCCAA	10140
GTAGCCGTAT CACTTCATCT ATCATTTTCG TAATCATTTT GATTATTACT GTATTGTTGA	10200
TTATTACTT TACAACCTAT GTTCAACAAG CAGAATACAA AATTCCAATC CAATATACTA	10260
AGGTGCACA AGGTGCTCCA TCTAGCTCTT ACCTTCCGTT AAAAGTAAAC CCTGCTGGAG	10320

545

TTATCCCTGT TATCTTTGCC AGTTCGATTA CTGCAGCCTG CGGCTATTCT TCAGTTTTTG	10380
AGTGCCACAG GTCATGATTG GGCTTGGGTA AGGGTAGCAC AAGAGATGTT GGCAACTACT	10440
TCTCCAACTG GTATTGCCAT GTATGCTTTG TTGATTATTC TCTTTACATT CTTCTATACG	10500
TTTGTACAGA TTAATCCTGA AAAAGCAGCA GAGACCTAC AAAAGAGTGG TGCCTATATC	10560
CATGGAGTTC GTCCTGGTAA AGGTACAGAA GAATATATGT CTAAACTTCT TCGTCGTCTT	10620
GCAACTGTTG GTTCCCTCTT CCTTGGTGTG ATTTCCATTT TACCGATTGC AGCTAAAGAT	10680
GTATTTGGTC TTTCTGATGT TGTTCCTTT GGTGGAACAA GTCTCTTGAT CATTATCTCT	10740
ACAGGTATCG AAGGAATCAA GCAATTGGAA GGTACCTAT TGAAACGTAA GTATGTTGGT	10800
TTCATGGACA GAACAGAATA AAAGTATTTA CTGAATCAGT AAATACTGAG GGAGTGAGG	10860
TTTAAACTCT GACATTTGTA AGAGTTGGAT CTCCCCTCTT CTATTTTGTT TTTAAATCGG	10920
GGTGAAAAGA CTTTTTGCTT CTATTTAAAA ATAAAATAAG GAGATCAAAT CATGAATCTT	10980
TTGATTATGG GCTTACCTGG TGCAGGTAAG GGAACCAAG CAGCAAAAAT CGTAGAACAA	11040
TTCCATGTTG CACATATCTC AACAGGTGAT ATGTTCCGCG CTGCAATGGC AAATCAAAT	11100
GAAATGGGTG TTCTTGCTAA GTCATATATT GACAAGGGTG AATTGGTTCC TGACGAAGTT	11160
ACAAATGGAA TCGTAAAAGA ACGCCTTTCA CAAGATGATA TTAAAGAAAC AGGATTCTTA	11220
TTGGATGGTT ACCCACGTAC AATTGAACAA GCTCATGCCT TGGACAAAAC ATTGGCTGAA	11280
CTTGGCATTG AACTAGAAGG TGTATCAAT ATTGAAGTA ACCCTGACAG CCTTTTGGAA	11340
CGTTTGAGTG GCGTATCAT CCACCGCGTA ACTGGAGAAA CTTTCCACAA GGTCTTTAAC	11400
CCACCAGTTG ACTATAAAGA AGAAGATTAC TACCAACGTG AAGATGATAA GCCTGAGACA	11460
GTAAACGTC GTTTGGATGT TAATATTGCT CAAGGAGAAC CAATCATTGC TCACTACCGT	11520
GCCAAAGGTT TGTTTCATGA CATCGAAGGT AATCAAGATA TCAATGATGT CTTCTCAGAT	11580
ATTGAAAAAG TATTGACAAA TTTGAAATAA AGCGTTTTTC AACTTGCAA AAATCCGCTA	11640
CAAATGTTAT ACTGAGATAG TCTGACTTAT AATTGTTGTC TCTGTGTCTA GAGGCATCGA	11700
ATCGAAATTT ATGGAGGTGC TTTTGCGTGG CAAAAGACGA TGTGATPGAA GTTGAAGGCA	11760
AAGTAGTTGA TACAATGCCG AATGCAATGT TTACGGTTGA ACTTGAAAAT GGACATCAGA	11820
TTTAGCAGG G	11831

(2) INFORMATION FOR SEQ ID NO: 66:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10726 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double

546

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CCCGGCATTT GAAAGCTATT CGTGAAGGAT TTATGATGGC AATGCCTTTG ATTTTAGTCCG	60
GCTCTTTATT TCTTATTCTA ATCAGTTGGC CTCAAGAGGC TTTTACAAAT TGGCTGAATA	120
GTGTTGGATT GCTAAGTATC TTGACAACTA TGAATCAGTC AACAGTAGCG ATTATCTCCT	180
TGGTCGCTTG TTTGCGTATT GCCTACAGGT TGTCGGAAGG ATATGGTACA GATGGTCCGT	240
CGGCAGGGAT CATAGCCTTA TCCAGTTTGG TATTGATGGC ACCTCGTTTT TCGAGTATGG	300
TTTATGATAA AAATGGGGAG CAGGTCAAGC AGTTATTTGG CGGCGCAATA CCATTTTCTA	360
GCCTGAATGC ATCTTCTTTG TTTATGGCGA TTACTATTGG ATTGGTTACA GCAGAGATTT	420
ATCGTATGTT TATCCAGCGC GGAATTACGA TAAAAATGCC AAGTGGTGTC CCAGATGTAG	480
TAAGTAAATC ATTTTCAGCT CTTTATCTG GTTTTACTAC TTTTGTTTTG TGGGCTTTGG	540
TCTTAAAAGG TCTTGAAGCG GCAGGAGTTG CAGGAGGTCT CAACGGACTC CTAGGTGCAA	600
TTGTGGAAC ACCGCTTAAG TTAATTGCAG GAACGCTTCC AGGTATGATT CTATGTGTTA	660
TTGTAAACTC ATTCTTTTGG TTCTGTGGAG TTAATGGGGG ACAAGTTTAA AATGCTTTTG	720
TAGACCCAGT TTGGTTACAA TTTACTACAG AAAACCAAGA AGCTGTGGCT GCAGGACAAA	780
CACTCCAACA CATTATTACA TTACCGTTA AAGATTTATT TGTATTATT GGTGGCGGTG	840
GAGCGACTAT TGGTCTTGGC ATTGTCTCT TCCTATTTAG TAAGAGTCGT GCGAATAAAA	900
CATTAGGTAA GCTAGCTATT ATACCGTCTA TTTTAAATAT CAATACAGCT ATTCTATTTA	960
CGTTTCCAAC AGTTTAAAT CCGATTATGC TGATTCCGTT TATTGCTACT CCTACAATCA	1020
ATGCCTTGAT TACCTATGTA TCAATGGCTG TAGGATTAGT ACCCTATACA ACAGGTGTAA	1080
TCCTTCCGTG GACAATGCCA CCGATTATAG GAGGCTTCCT TGCAACAGGG GCTAGTTGGC	1140
GAGGAGCTCT ATTACAAGTT GTTTTGATTT TGGTTTCTGT AGCAATTTAT TATCCATTCT	1200
TCAAAATTGC AGATAAACGC AATCTTGAAA AAGAAAAAGC TACTGTTGGA GGGAAATAAG	1260
ATGGTTATCA GAGTATTTGA TCAACAGAAA AATACTTATT CTAGCTTTGC CTTAGAGGAA	1320
TTAAGTTACT ATATGAATCG GGTCTTTAAG ACTAACATAG AGCTTGTCGA GGAGAAGGAA	1380
GCGGATATTT TTGTAGGATT AGTCAATAAA GAGGACAGAA AAGACCATGT TCTTATCTCA	1440
TTAGACAAGG GTAAGGGGAG AATTGAGTCT AATACAATTG TAGGTTTACT TATTGGAATT	1500
TACCGAATGT TTCATGAATT TGGGGTTGTG TATACTAGAC CAGGGCGCAG ACATGACTTT	1560
GTTCAGAGT TACGATTTGA AGATTTTTTA GATAAACAGC TATCTATAGA TGAAACAGCC	1620

547

AGTTACTATC ATAGGGGAGT ATGTATAGAG GGAGCGGATT CATTTGAAAA TATACTAGAT	1680
TTCAATTGATT GGCTACCTAA GATTGGGATG AACAGTTTTT TCATCCAGTT TGAAAAATCCT	1740
TACTCTTTTT TGAAACGTTG GTATGAACAT GAATTTAATC CATATCTAAA TAAAGAACAA	1800
TTTTCAAATG AATTAGTACA AGAATTGAGT GATAGGTTGG ATAAAGAATT GCAAAAAAGA	1860
GGTCTTATTC ATCATCGTGT TGGTCATGGA TGGACAGGTG AAGTTTTAGG TTACTCTTCA	1920
AAATTTGGCT GGAATCAGG TCTTAGTATT TCAGAGGAGA AGAAACCTA TGTCGCTGAA	1980
ATAAACGGGA AACGAGAAAT GTTTAATACG GCTCCGATTT TAACCAGCCT GGATTTTTC	2040
AATCCAGATG TAGCTGATAA GATGGTAGAA ATTATCAAGG ATTATGCCAA GAAAAGACCT	2100
GATGTTAACT ACTTACATGT ATGGTTGTCG GATGCTCGTA ATAATATTTG TGAATGCGAA	2160
AACTGTAGAC AAGAATTGGT TTCGGATCAG TATATTCGTA TTCTCAATCA ATTGGATAGG	2220
GCTTTAACGA GTGAGGGATT AGATACAAAG ATTTGTTTTT TGCTTTATCA TGAGTTGTTA	2280
TGGGCACCTC AGAAAGAAAA ATTAGATAAT CCTGAACGCT TTACCATGAT GTTGCACCG	2340
ATTACAAGAA CATTTGAAAT GAGTTATGCA GATGTAGATT TTGACAATTC CATACCTACG	2400
CCTAAACCTT ATATGCGTAA TAAAATTATA CTTCCGAATT CTCTTGAGGA AAATTTATCT	2460
TATCTTTTTG AGTGGCAAAA AGCATTTAAA GGAGATAGTT TCGTATATGA CTATCCTTTA	2520
GGGCGTGCTC ATTATGGCGA TTTAGGCTAT ATGAAAATTA GTCAAACTAT TTACAGAGAT	2580
GTATCTTATC TTTCCAACCT ACATTGGAAC GGTACATTT CGTGTCAAGA ATTACGTGCC	2640
GGATTCCCTC ATAATTTTCC TAATTATGTC ATGGGGGAAA TGCTCTGGAA GAAGACAAGA	2700
AGTTATGAAG AATTGATTGA AGAATACTTT TCTGCTTTGT ATGGGGAAAA TTGGCAGTCT	2760
GTGTTGAAT ATTTAGAAAA ATTATCCAAT TATTCCTCTT GTGATTATTT TAATGCAATT	2820
GGCAGCCGTC AAAGTGATGT TTTAGCGAAT CATTATTATA TAGCTTACAA TCTAGCTGAT	2880
AATTTTAC CAATTATTGA GGAAAATATT TCTAAGTTAT TAAATAGTCA AAAGGATGAA	2940
TGGAACACG TCAGTTATCA TCGTGAATAT GTTGTAAAGA TGCGGAAGGC TTTATATCTT	3000
CAAGCAACTG GAAAAACAAG GCAAGCTCAA GATGAATGGA GAAATGTGTT GAATTATATC	3060
CGTGGGCACG AATTGCTATT TCAATCTAAT TTGGATGTTT ATCGTGTAAT TGAAGTAGCA	3120
AAAAATTACG CTGTTTCCA CTTATAAATC ATAAGTATAG AAAATGAACT AAGGTATTCA	3180
GAGAAGATTG ATCCTAAATA TTATGAAAT TAAGGATTTT TAAGATATTT AGGGTCAACT	3240
TTCTATTTAT ATCGTAGCGA AGTCATTTTA ATAATGATGT GTAAAAGATG GATCAAGATT	3300
GAGGAGGAAG AAAGATGAAA TCAAAAGAAG AAATAAATAT GCTTGGTTTT ACAATTGTCG	3360

548

CTTACGCAGG AGATGCAAGG TCAGATTTGA TGGATGCTTT GGCCTTTGCG AGAGATGGAT	3420
ATTTTGAACA GGCAAGAGAA TTGGTTGAGT CTGCAAACGA CTCAATAGTG TCTGCCCATC	3480
GAGAACAGAC TAATTTATTA GCGGAGGAGG CATATGGAGA TAATTTTGAA GTGAGCTTTA	3540
TTATGATTCA TGGTCAAGAT ACTTTGATGA CAACGATGCT ATTGTATGAT CAGGTAAAGT	3600
TTTTTATTGA TGAATATGAA CGAATTCGAA AGATTGAAGA ACATATTGGT TTGCAATGAG	3660
GATTAGTCAT GGAAAATTTA CAGGTAAAG CCTTACCGAA GGAGTTTTTA TTAGGAACTG	3720
CTACCGCTGC TTATCAAGTA GAGGGTGCAA CTAGGGTAGA TGGCAAAGGA ATAAATATGT	3780
GGGATGTTTA TTGCAAGAA AATAGTCCGT TCTTACCAGA TCCAGCTAGT GATTTTTATT	3840
ATCGTTACGA AGAGGATATA GCTTTGGCGG CAGAACATGG TTTGCAGGCT TTGCGTTTAT	3900
CTATTTCTTG GGTTCGTATA TTTCTGATA TAGATGGGA TGCTAATGTA TTAGCTGTTT	3960
ATTATTACCA TAGAGTTTTT CAGTCTTGCT TAAAACATAA TGTGATTCCG TTTGTTCTT	4020
TACATCATTT TGATTCGCCT CAGAAAATGT TAGAAACAGG GGATTGGTTG AACAGAGAGA	4080
ATATTGATCG TTTCATACGA TATGCTCGCT TTTGTTTCCA AGAATTTACA GAAGTCAAGC	4140
ATTGGTTTAC AATCAATGAA CTGATGTCTC TTGCTGCAGG TCAATATATA GGAGTCTAGT	4200
TTCTCCAAA TCATCATTTT CAATTATCTG AAGCAATTCA AGCGAATCAT AATATGTTGT	4260
TGGCGCATGC TCTTGAGTC CTCGAATTTC ATCAATTAGG GATTGAGGGA AAGGTAGGTT	4320
GTATTCATGC TTAAAGCCA GGCTATCCTA TTGATGGGCA AAAAGAAAAT ATTTTGGCAG	4380
CTAAACGGTA TGATGTTTAT AATAATAAAT TTCTATTAGA TGGAACTTTT TTGGGCTACT	4440
ACAGTGAGGA CACGCTTTTT CACTTGAATC AAATATTGGA AGCTAATAAT TCTAGCTTTA	4500
TTATTGAAGA TGGTGATTTA GAAATTATGA AGAGAGCTGC ACCTCTTAAT ACGATGTTTG	4560
GGATGAATTA TTATCGTTCA GAATTTATTC GTGAATACAA AGGTGAAAAT AGACAAGAAT	4620
TTAATTC AAC AGGAATAAAA GGACAGTCTT CTTTAAATT AAATGCTCTA GGTGAATTTG	4680
TAAAAAACC TGGTATTCG ACAACAGATT GGGATTGGAA TATTTATCCT CAAGGGTTAT	4740
TTGATATGTT GCTTCGTATC AAAGAAGAAT ATCCTCAACA TCCGGTCATT TATTTAACTG	4800
AAAATGGTAC AGCCCTTAAA GAAGTTAAGC CAGAGGGCGA GAATGATATT ATTGATGACA	4860
GTAAGAGAAT CCGTTATATT GAGCAACATT TACACAAAGT TTTAGAGGCT CGAGATAGAG	4920
GAGTCAATAT TCAAGGCTAT TTTATATGGT CTTTGCAAGA TCAATTTTCT TGGGCGAATG	4980
GCTACAATAA GCGATATGGT CTTTCTTTTG TTGATTATGA AACACAGAAG AGATATATTA	5040
AGAAAAGTGC TCTTTGGGTA AAAGGGCTAA AACGGAATTA AGGTTAGCGA TTTGACTGAT	5100
GTTTAATATG TTTTAAATAT GAGGTGAAT TTTTATAGG AGGAGTTTTA TGGATAAGCT	5160

549

AGTCGCTGCC ATTGAAAAGC AACAAGGGAA ATTTGAAAAA ATTTCTACTA ATAACTATAT	5220
GATGGCTATT AAAGATGGAT TCATTGCTAC TATGCCTTTA ATTATGTTTT CAAGCTTTTT	5280
GATGATTATT ATTATGATTC CTAAAAATTT CGGAGTAGAG TTACCGAGTC CAGCTATTGT	5340
CTGGATGAGA AAAGTGATATA TGTAAACCAT GGGAGTTTGT GGTATTATTG TTTCAGGGAC	5400
TGTTGGAAAG TCATTAGTTG GAAATGTTAA CAGAAAAATG CCTCACGGAA AGGTAATAAA	5460
TGATATTTCT GCAATGTTGG CAGCCATATG TAGTTATCTG GTATTAACTG TAACGCTTGT	5520
AGTTGATGAG AAGACGGGAT CTACAAGTTT GTCGACAAAC TATTTAGGAT CTCAAGGATT	5580
GATAACTTCG TTTGTCACTG CCTTTATTAC TGTAATGTT TACCGATTCT GTATTAAGCG	5640
AGACATTACT ATTCATTTAC CTAAGGAAGT TCCTGGGGCT ATATCACAAG CTTTTAGAGA	5700
TATTTTCCCT TTTTCTTTTG TTTTACTTAT TAGTGGTTTG TTAGATATTG TATCTCGGTT	5760
TAGTTTAGAT GTTCCTTTTG CCCAAGTATT TCAACAAC TAAGTCTCTA TTTTAAAGGG	5820
GGCAGAATCA TATCCTGCTA TGATGTTGAT TTGGTTTATG TGTGCTTTCG TTTGGTTTGT	5880
TGGAATTCAT GGACCATCTA TTGCTTACC TGCTGTTACA GCTTTGCAAC TGAGCAATAT	5940
GGAAGAGAAT GCTCAACTTC TTGCAATGG GCAGTCCCT TATCATTCTT TAACACCTAA	6000
TTTCGGGAAT TATATCGCTG CTATTGGAGG AACGGGGGCT ACCTTTGTTG TACCATTTAT	6060
TTTGATTTTC TTTATGCGGT CTAACAATT AAAATCGGTA GGTAAGCTA CAATTACTCC	6120
TGTTTATT TTGCGTAAATG AACCTCTTCT ATTTGGTATG CCTGTTATTT TGAATCCCTA	6180
TCTTTTGTG CCTTTTGA TGAATCCCTG AGTGAATGTA TTTCTAGGAA AGGTCTTTAT	6240
TGATTTCTTT GGAATGAATG GATTTTATAT CCAGTTACCT TGGACCTTTC CTGGTCCCTT	6300
GGGATTGTTA ATTGGAACGA ATTTTCAACT TATCTCCTTT GTATTTTAT CTTTGATTTT	6360
AGTTGTCGAC ATATTGATTT ATTTGCCATT CTGTAGAGCG TATGATAGAC AGTTACTGGT	6420
GAAAGAAGAT ATGCAAGCT CAAATGATAT TATTTTAGAG GAGGATACAA GTGAATAAT	6480
TCCTGGTGAG ATAGATGAAA TAAAAAGTAA GGAGTTGAAA GTACTGGTTC TTTGTGCAGG	6540
GTCTGGAACA AGTGCAGCAAT TAGCCAATGC AATTAACGAG GGGGCTAACT TAACAGAGGT	6600
TAGAGTGATT GCGAATTCAG GAGCGTACCG AGCTCATTAT GATATTATGG GTGTTTATGA	6660
TTTAATTATT CTGGCCCCAC AAGTTCGGAG TTATTATAGA GAGATGAAGG TGGATGCAGA	6720
AAGATTAGGT ATTCAGATAG TTGCTACCAG AGGAATGGAA TATATTCATT TAACAAAGAG	6780
TCCAAGTAAA GCCTTACAAT TTGTATTGGA GCATTACCAA GCTGTGTAGT AAGTTTTC	6840
ATCTTTTATT TGAGTAAAGA TTTGTTTAC AGATAGGCTT GGATTTAAAA ACGTTCCCCC	6900

550

TTTTTTAATA TAAGAATCCC TCTTTCACAA TTGTAAAAAG AGGGATTTTG TATTTTATCT	6960
CTTAGACCAA GTTCTCTTCA TAAAGAGAAG GAGGATTGGG TAAATCTCCA AGCGCCCTGC	7020
AATCATTGCA AAGGATAGGA GAATTTTGA GATGGGACTA AAGATTGAGA AACTAGAAGT	7080
GGTTCCTAGA ATAGGCCCGA TATTATTGAA ACAGCTAAAG ACAGCGCTGG TCACGACCAG	7140
AAAATCATTG CTATCTAGGC TGACAATAAA GATAAGCGCT AGCAAAATCA TAGCATAGAT	7200
GACAAAGTAC TTGAGAATCT TATGCTGGGT ATCTTTGTCA ATCACCCTTT TATTAACATG	7260
GAGGGTCAAA ACACGGTGGG GCGATAGGAT TGACAAAATT TGGTTTTTGG CAATTTTGA	7320
AAGGATGAGG CCTCGAATAA TCTTGAGTCC ACCTGCAGTT GATCCAGCAG AGCCACCGAT	7380
TGCCATGAGG AAAAGGAGGA TAAACTGGGA GAAGAGGGGC CAGTTGGTAA TATCTCCATA	7440
TCCAAAACCA GTTGTTGTAA TGATGTTGGA AACCTGGAAG AAGGTCATTT CAAAGCTCTT	7500
TGAAAACCTT GGGTAGAGGT AGAGGGTGTT GAGGCTAATC AAGCCTGTAG AAACCAGTAC	7560
AATGACCAAG TAAGCCCTAA GCTCTTCATC TCCAAAGAAG GCCTTGATGC GACGGAGCAT	7620
GAGTAGTAG TAGAGGTTGA AATTTACTCC AAAAACCAGA ACTCCGATAC TGACCAGATA	7680
GGTAATCACT GAGCTGCCAT AGTGGGCAAT TCCGTCGTTA TAGACGGTAA AGCCTCCAGT	7740
TCCCGCTGTC CCCATAGCAA TAACAAAACCT ATCGTAGAGA GGCATACCGG CTAGATAATA	7800
GATGATGACA AAGAGGGAGA AGAGAGCTAG ATAAAGGAGA TAGAGAATCT GGGCAGTGTT	7860
TTTGTAGTTG GATACAACCT TGCCAAAAC AGGACCTGGA ACCTCAGCCT TCATCACCTC	7920
TAGGTGGCTA TTTTGGCAT TGTCCATAAT AGCAAGTGCA AAAACAAGCA CTCCCATCCC	7980
TCCAATCAAG TGGGTAAAAC TTCGCCAGAA GAGGAGGGAA CGGCTGAGAA CCGAAACGTC	8040
GTTCAAAATA CTTGCTCCAG TAGTTGTAAA TCCAGAACTA ATTTCAAAAA AGGCATCAAT	8100
AAGGCTGGGG ATTTGCCAG AAAAGACAAA GGGGAGACCA CCAAAGAAAG ACCAAAGGAT	8160
CCAACAGAGG GCAACGATCA AGACTCCCTC CTTGGCATAA ATCCGTTGAT TTTTGGCTT	8220
CTGTAAACTC CCTGAACCGC CTAACAATAC GAGAATCCCT ATGGTCGAAA AGAGGGCTGT	8280
AAAGACTTGG CTCGATTAC GGTAATAGAC AGCAATCGCA ACAGGAACCA AAAGAAGAAC	8340
AGCTTCAATC AAAAGTAATT TTGAAAGGAG GTAACGAATC ATACTTTTAT TCATTTCTTA	8400
CCTCGCATC AAGTCATAAA TCTTGGTGAT GTTGGCAAC AAGGTTGTTA CTAGGAGCTT	8460
GTCTCCAACT TCCAACATAT CCTCCCAGT TGGGAAAATA GTCTTGCCCT TTCGAATAAT	8520
GGCTGCAATA AGAACCCCTT TTTTCAATTT CAGTTGAGAA AGAGGTTTGG CAGTCATTTT	8580
ATTGGCTTCC TTGATATGGA ATTGCAGGGT TTCGATTGG CCATTGGCTA GATGGTGCAT	8640
AGCTTGAAGG TCTGAATACT GGGCATTAAC TCGACCACGA ATAAAGTGCA TAATCGTATC	8700

551

TACAGCGATG CTTTTAGGTG TGATGATACT TGAAAAATCA GGCGCATTGA TAATCTCGAG	8760
GAGACTGGTA CGATTGACCT TAGTAATATT TTTCTGTACA CCTACCCTGT CAAGGAACAT	8820
AGATGTAATC AGATTTTCCT CATCGACTCC TGTTAGAGTC GCAACGGCAT CATAGTGTTC	8880
AGCACTTTCT TCCAGCAGGA TATCTTTTGC GGTTCATCT CCTTGAACGA TGTAGAGATT	8940
TGGGAATTTT TCGCTAAAGA AGCTGGCGAT TTCAGGATTG ATTTCAATGA CTTTGTATC	9000
GATACGACTA TCTTTGAGAA TACCAAGTAG ATAATAGGCA ATTCTACCTG CCCCACCGAT	9060
GAGAAGGCTC TTCACGGCGC GTGATTTAAA ATAATTATGG AAGAGTATCA TATCGACACG	9120
GTTACCAGTG ACAAAGATTC TATCTTTATC CTGTACAGTC ATGTCACCGC TTGGAATGAT	9180
AATTTGATGA TCCCTCTCTA TCGCACAGAC AATGACATTA CCAAATTTT TACGAAAATC	9240
AGAAATGGGC ATTTGGCAAA GACCGTGGT GGACTTGACG ACAAATCCA TGAGGCTAAC	9300
GCGTCCACCA GCAAAGCGTT CGACAGACAG GCGTTGGGG AAGTCAATGA TATTCGCGAT	9360
AGCGCGGGCA GCCAAGAGCT CAGGATTAAC GATAAGAGAA AAACCGAGAA TATTCTTTTC	9420
CTTGAAATAA GAGTTAGAAT ATTCAGGGTT CCGCACCCGA ACGATAGTTT CTTTAGCTCC	9480
CATTTTCTTG GCTAGAACTG CTGCAATCAT GTTGACTTCA TCGTGCTCAG TCAGGGCGAT	9540
AAAGATATCA CAATCTTGA CGCTGGCTTG CTCAAGAATG GCAAAATCGG CCCCGTTACC	9600
AAGGATACCA ATGATATCAA AGCGACTGAC AATATGATTG AGAACAGCTT CGTCTTGCTC	9660
AATCAGCAAA ACATCATGCT TTTCTGCAAC CAAGGAGCGA CAGAGGGCAA AACCACCTTT	9720
TCCCCCTCCG ACAAGGATAA TTTTCATAAT AAAACCTACT TTTTCATGAT GTAACATCA	9780
TACCCTTTTT CAAGAAAAAA TGCACCTACT AGCTAATAAC AAGAGTTTTT AGTGAAAATT	9840
CGCTATAAGG TAAAACTATA CCCTAACCAA TTGAAATAGC TATTAGCGAC TTTCTCTGAA	9900
ATATGGTATG ATAAAGGATA TACAAGGAGA TAAATGAAT AATAATTTAC TGGTATTACA	9960
ATCAGACTTT GGTCTGGTTG ATGGTGCGGT ATCGGCTATG ATTGGAGTGG CTTTAGAAGA	10020
GTCTCCAACC TTAAAAATAC ATCACTTGAC GCACGATATC ACGCCTTATA ATATTTTGA	10080
GGGGAGCTAT CGTCTCTTTC AGACGGTGA TTAAGAGTGA GTTGCCAAGA CTGCAAAAAA	10140
GGTTGTCGAT CCAGGTGTCG GTTCGAAACG TAAGAGTGA GTTGCCAAGA CTGCAAAAAA	10200
TCAATACATT GTCACGCCAG ATAATGGGAC GCTTTCCTTT ATCAAGAAAC ACGTTGGCAT	10260
TGTAGCCATT CGTGAGATTT CTGAGGTGGC CAATAGGCGT CAAAACACAG AGCATCTTA	10320
TACCTTCCAC GGTGCGTATG TCTATGCCTA TACTGGTGCT AACTGGCCA GTGGTCACAT	10380
TACTTTTGAG GAAGTAGGGC CAGAGCTCAG TGTGGAACAG ATTGTAGAGC TTCCAGTCGT	10440

552

AGCGACCATC ATAGAAGATC ATCTGGTGAA GGGAGCCATT GATATTCTGG ATGTGCGTTT	10500
CGGTTTCGCTT TGGACCTCTA TCACACGGGA AGAATTTTAC AAGCTGGAAC CAGAATTTGG	10560
TGATCGTTTT GAAGTGACCA TCTATCATGC TGATATGCTG GTCTATCAAA ATCAGGTTGT	10620
CTATGGCAAA TCATTTCAG ATGTGAGAAT TGGGCAACCs ATcTTTACrc TCAGCaTCTt	10680
CGATTAGCTG GGCAATTCGT TCTAGTTGGA TTTCGTCAAT CAAGGT	10726

(2) INFORMATION FOR SEQ ID NO: 67:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7163 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTATCTTTAA CGATATCAAT CAAGATCTGG TCAATAAAGG GATTGGGGCT TATCGTGAAG	60
TTGGCATCCA AGCCCATGGA TATGTCTGTG ACGTGACAGA CGAGGACGGT ATCCAAGCCA	120
TGGTCAAGCA AATCGAACA GAGGTTGGTG TCATTGACAT CCTCGTTAAT AACGCTGGTA	180
TTATCCGCCG AGTTCCAATG TGCGAAATGA GCGCCGCTGA TTTCGTAAG GTCATCGATA	240
TTGACTTAAA CGCACCATTT ATCGTTTCAA AGGCAGTTAT TCCTTCTATG ATAAAGAAAG	300
GCCATGGAAA GATTATCAAT ATTTGTTTGA TGATGAGCGA ACTGGGACGT GAAACAGTTA	360
GCGCTTATGC TGCTGCTAAA GGGGGCTTGA AAATGTTGAC CCGCAACATT GCGTCTGAAT	420
ACGGTGGAGC CAATATCCAA TGTAACGGAA TTGGACCGGG TTATATTGCC ACTCCTCAAA	480
CAGCACCTCT TCGTGAATTG CAAGAAGATG GTTCTCGCCA CCCATTTGAC CAGTTCATCA	540
TTGCAAAAAC ACCTGCTGCA CGTTGGGGAA TACTGAAGA TTTGATGGGC CCTGCTGTCT	600
TTCTCGCTAG TGATGCCAGC AATTTTGTCA ATGGCCACAT CCTATATGTA GATGGCGGTA	660
TCTTAGCCTA CATCGGAAAA CAACCTGAGT AAAAATAGAA AGAAGATCTT ATGAAAATCG	720
CATTAATCAA TGAAAATAGT CAAGCTAGCA AGAATCACAT TATTTACGAT AGTCTAAAAG	780
AAGCGACAGA TAAAAAAGGC TACCAATTAT TTAACATATGG TATGCGTGGA GAAGAAGGAG	840
AAAGTCAATT AACTTATGTG CAGAACGGAC TAATGGCTGC CATCCTTTTA AATACAAAGG	900
CAGTTGACTT TGTGTTACC GGCTGTGGTA CGGGTGTAGG GGCTATGCTT GCTTTAAACA	960
GCTTCCCTGG TGTGTCTGT GGTCTAGCAG TGGACCCAAC TGACGCTTAC CTTTATTCTC	1020
AAATCAATGG TGGTAACGCC TTGTCTATCC CTTATGCCAA AGGATTTGGC TGGGGGGCAG	1080
AACTGACCCT CAAATTGATG TTTGAACGCT TATTTGCTGA AGAAATGGGC GGTGGCTACC	1140

553

CAAGAGAACG TGTAATCCCT GAACAACGCA ACGCTCGTAT CTTAAACGAG GTGAAACAAA	1200
TCACCCACAA TGATTGTAGT ACCATCCTTA AAATAATCGA CCAAGACTTC CTCAAAGACA	1260
CCATCTCTGG CAAATACTTC CAAGAATACT TCTTTGAAAA CTGCCAAGAT GATGAAGTTG	1320
CTGCTTATTT GAAAGAAGTA TTAGCCAAGT AAAGCTATTC TAAACCAGAA AGGAACTAAT	1380
GGATGACGAA AATATTACTG TTTGGCGAAC CATTAATTCG AATTTACCA TTAGATGCCA	1440
CCAGTATCGG CGATCATGTT GCCAGTTCGA CTTATTTTGG CGGATCAGAA ATTAACATCG	1500
CTTGTAATTT GCAAGCCCTG GGTATCTCAA CGAAAGTTT TACCGCACTC CCTGCCAACG	1560
AGATTGGAGA TCGTTTTCTC ACATTCTTGA AACAGCACCA AATCGATACC AGTTCAATCT	1620
GTCGGCTTGG CGATCGAATC GGCCTCTACT ATTTGGAGAA CGGCTTTGGT TGTGTCAAA	1680
GTGAAGTTTT CTACGATCGT AAGCATACGA GTATCAGCCA GATTCGGCCA AACATGCTAG	1740
ATATGGATTC TCTCTTTCAG GGGATTAGCC ATTTTCATTT TAGTGAATC ACCGTAGCTA	1800
TGCGTCAAGA GGTCCGTGCG ATCCTTCTCC TACTCTTGA AGAAGCCAAG CGCCGAGGAA	1860
TTGTGCTTTC AATGGATCTC AATCTGAGAA CAAAGATGAT TTCAGTCTTA GAAGCCAAGT	1920
ATGAATTTTC TAAGTTTGCA CGTTTACTG ACTATTGCTT CGGTATTGAT CCTCTCATGA	1980
TTGATGACCA AAATCTAGAG ATGTTTCCAA GAGACAGTGC TAGCCTAGAA GAGGTGAAA	2040
ATCGCATGCG ACTTTTAAAA GAAGCCTATG GTTCAAGGC CATTTTCCAT ACCCTCCGCT	2100
CTAGTGATGA GCAAGACAAA AATGTCTATC AAGCCTATGC TCTAGAAGAA CTATTTGAAG	2160
AGTCTGTCCA ACTAAAACT GCAGTCTATC AACGAATTGG TAGCGGGGAT GCCTTTATAT	2220
CTGGTGCCCT TTACCAACTA CTCCATCATT CCTCCCTAAA AACTACCATT GACTTTGCAG	2280
TTGCGAGCGC AACTCTCAA TGCACCTTC CAGGAGACCA TCTCTCCACT TCCTCAACTA	2340
GTATTGAAAA TTTACTGGCA AATGCACAAG ATATCATTCG TTAGGAGAAT TACATGACCA	2400
AATCAGATAC GATTATTGAA CTAATAAAAC AAAAAATTGT CGCTGTTATT CGAGGAAATA	2460
CAAAGGAAGA AGGACTACAA GCCTCGATTG CTGTATCAA GGGCGGTATC AAAGCTATTG	2520
AAATCGCCTA TACCAATCAG TATGCAGGAC AAATCATCAA GGAACCTGTA GACTTGTATC	2580
AGGACGATCA GAGTGTGTTG ATCGGTGCAG GACTGTGCT TGATGCCGTA ACTGCTAGAG	2640
ATGCCATTCT AGCTGGAGCA AATTACGTG TTTCTCCATC TTTCCATGCT GAAACTGCCA	2700
AAATGTGCAA TCTCTACAGC ACACCGTACA TTCCAGGCTG TATTACCCTC ACAGAGATCA	2760
CGACTGCACT TGAAGCCGGT AGTGAAATCA TCAAACCTCT CCCAGGTAGT ACTCTCAGTC	2820
CAGCATATAT CTCTGCAGTC AAGGCACCGA TCCCACAAGT TTCCGTAATG GTAACCGGAG	2880

554

GAGTCGGCCT AAACAACATC CCTCAATGGT TCGCTGCTGG TGCAGATGCC GTTGAATTG	2940
GTGGCGAACT CAATAAACTC GCTTCCCAAG GCAACTTTGA CCGCATCAGC GAGATTGCCC	3000
AACAGTATAT TACACTCAGA TAAAATCATA ACTACCCGTC TAACGGGTGG TTTATCTCAG	3060
AGCTATAAGC CCAAATCATC AGCCAGCGCC TAAAGACGCT GGCTTTCACG TTGTTCAAGC	3120
CTTATTGCTC TTGACTCGTC ACTTGCCTCT TTAAGAGACT TTGGTATTAC TTACCACTAT	3180
CCCTAAAGGG ATCCTCATAT TCTTTTACAC TCAATTTATC TAGTGCTATA GTAGATTGAA	3240
ACTGGAATAG TACACCTCTG CTTCTAAAAC ATGTGTTAAA ATCGATTGA CTGTCCTGAT	3300
CGATTTTGTC CTGTTCTTAT TTCATTTTAC TATATATCAT ACTTTACTCG TTCTCAAATT	3360
TTTCATACTCA TGAAGAAATC ATCCACTCGA TAATTTCTTT AATCTTGACT ATATTTCTTA	3420
ATTGTGGCTT CATTAAGCCC TACTGGACTT ACATAATAAC CTTCTCCCA GAAATGCCGA	3480
TTCCCAAAC TGTACTTGAG ATTGGCGTGT TTGTCAAACA TCATGAGTGC ACTTTTGCCT	3540
TTTAAATACC CCATAAACT TGAAACACTT AGCCTCGACG GAATACTGAC TAACATGTGT	3600
ACATGGTCTG GCATTAAGTG ACCCTCGATC ATTTCAACAC CTTTATAACT ACACAAGCGA	3660
TGAAATATTT CGTCTAACT ACTTCTATAT TGATTATAGA TGACTTTTCG TCTATACTTA	3720
GGGGTGAACA CAATATGATA GAACACCTCC ACTTTGTGTA TGATAAACTA TGAGTCTTTT	3780
GTGCCATATT TTTTCTCCTT TCGCTTTACA ATTGGATTGA ACACCTTTAT TGTATCGCGT	3840
TTGGAGTTT TTTGGTATAA CCTTCGACGC GCACCCGTAT AGCGGGTGGT TGTTTTGTCT	3900
CGCACCTCAC GGAGCGAGAC GGACTAATAT AGTGGAGTGA AATAGGATAC GAACAAATTG	3960
ATTAGGAAAA TCAAATGAAT TTATAGAAAT CTTTTCAGC TTATAACGTT CTATTCTAGT	4020
TTCAAAACGC TATAGTCACA TAATAATGAA GTAAAAAAGG ATAAGTATCA ACTTATCCTT	4080
TTTTAAAGA AAAATCCGAA GATATTTGGC CTTCTTCGGA TTTTTCCTAT TTTCCACAGT	4140
TTCATGTAAT TCATCTAGAT GATGAACAAA TTAGTTGTTT TTTCTCTAC GGAATAGATA	4200
AAATGCCCCA AGTAGCAAGA ACCCTAGACT TGCCAAGATT GACTGACCTT CTCTGTCTG	4260
AGGGAGATTC TTTTGATCCG AATGGTTCTT TTCCTCTTCA GATTTTTCCT TTTCTTTGA	4320
ATTCTGTACT TGTGGCTGAG CTGCTTGCTC TAGCTTTTTA AAGACTTCCT GATCTGGAGC	4380
TGATTCCTGG GTTTCAGGAT TATAGTAGGC AATCTTATAT TCATCCCCTT CTTTTCGAAT	4440
GGTATAGACT CCACGTTTCA AACTTTGGAA TTGGTTGGAA ATAGTAGAGA CAGAATCATC	4500
ATATTTTACA ATGCCCCAAA CTCCTTGTTT AGCATCATAA ACAGACTGAA GGGTTTCGTT	4560
ATTTTCGATG AGGCTACTTT CTAACCTCTT TATCATTTGA TTGAAGGTGG CACGATCCAC	4620
GTTAGGAATG AGCATATAGC CATAAGAATC TCTATTTTGC TTATGAGCCT GACTAATCGT	4680

555

AAGAAATCA TTTTCAACTT CCTTGTCTGA CTGTCCTTCA TTGATATCCT TCCAGGCTCC	4740
CTTTTGCAAA GCCTTACTCA TACTGATTGA ACTCTTCTTA AAGAAAAAGT AACCAATATT	4800
CTTTTTCGAA TCGAACGATT CTAAAAAGAC ACTTTGGGTT TCAGGATAAT CCTTTTCTTG	4860
TTCTGTAAGG GAGGCTTCTT TATCATTGAC ATAGACTTTA TATGGATTAC CTGATTCCAG	4920
TTTCTCTGG TCAATTGTAG TTGCAGCAGT ATCTGTTGAA GTGTTTTGGA TATTGCTTCC	4980
TAAAAAGGCG ATCTTATCCT TTAGCATAAA CCAGCTCTTA TGAGCAGTCA ATGTTTGATT	5040
CCAGTTGGTG AAATCCATGG TTGCTGTCGC ATTGGCATCA TCTAGTTTGC TCGTTCCAAC	5100
GAAAGCAGAC GGTAAACTT TACCTGTATC GCTATCCGCT CTCTTAGCAT CCGTCTCTGT	5160
TGTACCAGGC ATCTTATATG GATTAAGTGT TGGCCAGTAG CCATCGCTAT AGTGACTCAA	5220
ATCGCCATTG TAAAGATAGA ACATCCCATC ACTCGTATAC CAACCACGTT TATTTTCTT	5280
GTTCATGTGT TCGTAATTCA AGGTACGACT GGAAAAGAGT GACAAGCCAA ATCCAAACCC	5340
TTTCTCTGCA TTGTACATGG CTGTTTATC CATCTTGTTA AAGGCAGATA GGTAACCTGG	5400
TCTTGGAACA CTTGCGACTC CTGCATCACT TAACAAGGAT TGCATCAAAC TGATATCCTT	5460
ATAAGTCTTC AAATTCTTAA AGACATCATA ATAAGTATCC GATTGAACAA TGGTCTTCAC	5520
AAGACTCTGC AAACATTGTT TGGTTTCTCC TTCAGACATA TCCGCTATTG GGTGAATCCC	5580
TCTTAGTACT TCTACTGCGG CCACGTGCCC CTCGCTATTT GCACGACTGA TCGAGCGTCC	5640
ACGACTCATA TCCATCAACT CTCCATTAC CAGCAAAGGA GCAAACGATT TATCAATCCA	5700
GTGGTACATG GTTTCATTT TATCTTTATC GATTGGATTG TTGGTCTTTT GAATGACTGG	5760
CAACAGTTGA GACAGGCCAT CAATCAAAC ATTCCCATAA GCACCCGTAT AGGCAACATT	5820
GGTGTGGTCG ATATAGGATC CATCTGATA AAAACCTTCA CCTTGGTCTA CCAACTTGAA	5880
CACTTGCTCA ATCGAGCGAA TGGTAGAAGA AATTTCTTGA TCATCCTTAC GCAGTAAACC	5940
AGCTATTACT TTTACCTTTC CCATATCAAC TAAGTTTCCA CCTAGAGCCT TGAATGGGTT	6000
ATCAGTCGTC TTTTCGAAAT GTTCGGGATC TGGTACAAAT TTTTCAATCA CATCTGTATA	6060
TTTTTTAAT TCCTCATCAG AGAAGTATTC TTTCATCAGA GACAAGGTAT TGTGATGGC	6120
ACGAGGTGTA CCGATTTCAT AATCCCACCA GTTCCCAACA ATGCTCTTTT CACTATTGTA	6180
GACATGTTTA TGCATCCATT CCATGGAATC CCTGACTGTT CGAACGACAG TTTCATCTTG	6240
ATAATAACGA GAAGAAGGAT TGGTCACTTG CTGGCCATC TCCTCCAATT TCCGATAAGT	6300
GGCAGTCAGA TTTGCAGACG TTTTATAATT TGAAAATTTT TCCCACAAAT AGGTGCGGTC	6360
CGCCTGACTT GAAATACTGG ATAGGCTATC AGCTACCTTT CCTTCCAATT CCTGGTTTAA	6420

556

TTTGCCATC TGTTCATTT TAGAATCATA GTATTGATTG CCAGCGATGA TGCCATTCCA	6480
GTCATCCAAA CGGTCTGTGT ATGCATCCTT AACAGAGGCC AGAATCTTCA AAGGAATCTT	6540
TTTCACTTCC TTGCCATCTT TACTGACAAT GACATTGGTT GTCCCTTCCT TAAGAGGTTT	6600
TAAAATTCCA TTTTGTACTG AAGCAACGTC AGGATTTTCT ACCTTATAAG TATAGTCCGC	6660
AAGAGAAAAA ACATGTTTTT TTCCAATTGG TAAATCAATC TTTTCCTCAA GCTGTTTATC	6720
TGTTTGAGAA TCCTCAGAAA GCTGGTCTGC TACCTCTACC AGCTCAATAT CCTTAAAGGA	6780
AACAGTCCCA GTTCCTGTTT CATAGAATAA CTCCAGCTTG ATTTTATCAA CATCTAAAGT	6840
CGGGCTATAG TCTGCTTCAA TGGTCTGCCA GTCCTTGTGTT CCTGACGTCG TTGCAGAAAT	6900
CCACAATCGC TTGTCCTTAC CACTTTCCTC AATGATACGA ACTTTGGCAA TCCCGATTTT	6960
ATTATCTGTT TTAATCTTGA AACGCAGTTT ATACTTTTTC TTAGCTTCAA TAGGAACCAT	7020
ACGGTGAAGC GCTGCCCTTA ATTTCTCATG GCTTGAGATA GTGATAGCCC CATCCTTAGC	7080
CTCAATGACT CGAGTTGAGG CATCTGCACT ATTCTTCTGG TCTACCCAAG CTGACCACCC	7140
CCTGAGCTTT GCTTCCTGTC CGG	7163

(2) INFORMATION FOR SEQ ID NO: 68:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9244 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

CGTTATAACA TACATGTAAG CGGTACCCAA AATGGTGCCA AGTCAAAATT TTAAAGGAGG	60
AAAATACATG TCTTCACATC CAATTCAGGT CTTCTCAGAA ATTGGGAAAC TGAAAAAAGT	120
TATGTTGCAC CGTCCAGGCA AGGAGTTAGA AAACCTGTTG CCGGACTATC TTGAAAGGCT	180
TCTTTTGTGAT GATATTCCTT TCTTGGAAGA TGCTCAAAAA GAACATGATG CATTTGCCCA	240
AGCTCTTCGC GATGAAGGAA TTGAGGTTCT CTACCTAGAA CAACTCGCTG CTGAATCATT	300
GACCTCTCCA GAAATCCGCG ATCAATTATG CGAGGAATAC TTAGACGAAG CCAACATCCG	360
TGATCGTCAA ACCAAGGTTG CTATTCGTGA ATTGCTTCAC GGCATCAAGG ACAACCAAGA	420
ATTGGTTGAA AAAACAATGG CTGGGATTCA AAAAGTTGAA TTGCCAGAAA TTCCTGACGA	480
AGCTAAAGAT CTAAGTACT TAGTTGAATC AGAGTATCCA TTGCAATTG ACCCGATGCC	540
AAACCTCTAT TTCCTCGCG ACCCATTTGC AACAAATTGGA AACGCCGTAT CGCTTAACCA	600
CATGTTTGCA GAACTCGTA ACCGTGAAAC ACTCTACGGT AAGTATATCT TCAAATACCA	660

557

CCCAATCTAT	GGCGGAAAAG	TGGATTTGGT	CTACAACCGT	GAAGAAGATA	CGCGTATCGA	720
AGGTGGAGAC	GAGTTAGTTC	TTTCTAAAGA	CGTCCTTGCA	GTAGGTATCT	CTCAACGTAC	780
AGACGCAGCT	TCTATCGAAA	AACTTTTGGT	CAACATCTTC	AAGAAAAATG	TTGGCTTCAA	840
GAAAGTTTTC	GCCTTTGAAT	TTGCTAACAA	CCGTAAATTC	ATGCACCTGG	ATACTGTCTT	900
CACTATGGTA	GACTATGACA	AGTTCACCTAT	TCACCCAGAA	ATCGAAGGCG	ACCTTCACGT	960
TTACTCAGTT	ACTTACGAAA	ACGAAAAACT	TAAAAATCGTT	GAAGAGAAAAG	GTGACTTAGC	1020
TGAACTTCTT	GCTCAAAACC	TTGGTGTAGA	AAAAGTTCAT	TTGATTTCGTT	CGGGTGGTGG	1080
CAATATCGTA	GCAGCTGCGC	GTGAACAATG	GAACGACGGT	TCTAACACTT	TGACCATCGC	1140
ACCTGGTGTG	GTAGTTGTTT	ATGACCGCAA	TACCGTGACC	AATAAGATTT	TGGAAGAATA	1200
CGGGCTTCGC	TTGATTAAAG	TTGCGGAAG	TGAATTGGTT	CGGGGCCGTG	GTGGACCTCG	1260
TTGTATGTCT	ATGCCATTTG	AACGTGAAGA	AGTGTAATCG	CTGTTTCGATA	TTCGTCAATA	1320
GAAAATGTAA	AAAATAGAAA	GAGGAAATAA	TAAAATGACA	AATTCAGTAT	TCCAAGGACG	1380
CAGCTTCTTA	GCAGAAAAAG	ACTTTACCCG	TGCAGAGTTA	GAATACCTTA	TTGGTCTTTC	1440
AGCTCACTTG	AAAGATTTGA	AAAAACGCAA	TATTCAACAC	CACTACCTTG	CTGGCAAGAA	1500
TATCGCTCTC	CTATTTGAAA	AAACATCTAC	TCGTACTCGT	GCAGCCTTTA	CAACTGCGGC	1560
TATCGACCTT	GGTGCTCACC	CAGAATACCT	CGGAGCAAAT	GATATTTCAGT	TGGGTAAAAA	1620
AGAATCTACT	GAAGATACTG	CTAAAGTATT	GGGACGTATG	TTTGACGGGA	TTGAATTCCG	1680
CGGATTCAGC	CAACGTATGG	TTGAAGAATT	GGCAGAATTC	TCAGGCGTTC	CAGTATGGAA	1740
CGGTCTAACT	GACGAATGGC	ACCCAATCA	AATGCTCGCT	GACTACTTGA	CTGTTCAAGA	1800
AAACTTCGGT	CGCTTGGAAG	GCTTGACATT	GGTATACTGT	GGTGATGGAC	GTAACAACGT	1860
TGCCAACAGC	TTGCTCGTAA	CAGGTGCTAT	CCTTGGTGTC	AATGTTTACA	TCTTCTCACC	1920
AAAAGAATCT	TTCCCAGAAA	AAGAAATCGT	TGAATTGGCA	GAAGGATTTG	CTAAAGAAAG	1980
TGGCGCACAT	GTTCTCATCA	CTGAAGATGC	TGATGAAGCA	GTTAAAGATG	CAGACGTTCT	2040
TTACACAGAC	GTTTGGGTAT	CAATGGGTGA	AGAAGACAAA	TTGCGAGAAC	GTGTAGCTCT	2100
TCTTAAACCT	TACCAAGTCA	ATATGGACTT	AGTTAAAAAA	GCAGGCAATG	AAAACCTGAT	2160
CTTCCTACAC	TGCTTGCCAG	CATTCCACGA	TACTCACACT	GTTTATGGTA	AAGACGTTGC	2220
TGAAAAATTT	GGTGTAGAAG	AAATGGAAGT	AACAGACGAA	GTCTTCCGCA	GCAAGTACGC	2280
TCGCCACTTC	GATCAAGCAG	AAAACCGTAT	GCACACTATC	AAAGCTGTTA	TGGCTGCTAC	2340
ACTTGGTAAC	CTTTATATTC	CTAAAGTATA	ATTTTAGATA	ATAAACCGTC	TACCAACAGC	2400

558

TATGAGGGCT GCGACTAATA GCTTTAGTCC GGTCTCTTTT TATGTAATGG TAATCTATTA	2460
TTTCTTATAA AATATGTGAA AAATCATTA AATTGAAATCT AAACGCATTC TATTGAGTGT	2520
GATAAAGGAG AATTATGGC AAATCGTAAA ATTGTAGTAG CTTTGGGAGG AAATGCGATT	2580
CTTCTTCTG ACCCATCAGC AAAGGCTCAA CAAGAAGCTT TAGTTGAAAC AGCTAAGCAT	2640
CTTGTA AAAAT TGATTAAAAA TGGAGATGAT CTGATTATCA CTCACGGTAA TGGACCTCAA	2700
GTTGGGAATC TCTTGCTCCA ACATTTGGCA TCAGACTCTG AAAAGAACCC TGCCTTCCCA	2760
CTCGACTCAC TTGTCGCTAT GACAGAAGGT AGCATCGGTT TCTGGTTGAA AAATGCTTTG	2820
CAAAATGCTC TCTTGGATGA AGGCATCGAA AAAAATGTTG CCTCTGTTGT AACGCAAGTT	2880
GTCGTAGATA AAAATGATCC AGCTTTTGT AACTTGAGTA AACCAATCGG TCCTTTCTAT	2940
TCAGAAGAAG AAGCAAAAGC AGAAGCCGAA AAAAGCGGAG CGACTTTCAA GGAAGATGCT	3000
GGCCGTGGCT GCGTAAGGT CGTTGCCTCA CCAAAACCTG TTGACATCAA AGAAATTGAA	3060
ACCATCCGTA CTCCTTTAAA TAATGGTCAA GTCGTCGTAG CTGCAGGTGG TGGCGGTATT	3120
CCCGTCGTCA AAGAAAACAA TGGACATTTG ACTGGTGTG AAGCGGTTAT TGATAAAGAC	3180
TTGCGCTTCCC AACGTTTGGC AGAATTGGTT GATGCAGACC TCTTCATCGT TTTGACAGGT	3240
GTAGATTATG TATTGTGTTAA CTACAACAAG CCAAACCAGG AAAAATGGA ACATGTGAAT	3300
GTTGCCCAGC TGGAAGAATA TATCAACAA GATCAGTTG CACCAGGTAG CATGCTTCCA	3360
AAAGTAGAAG CAGCTATCGC TTTTGTCAAT GGTGCTCCAG AAGGAAAAGC AGTTATTACT	3420
TCCCTTGAAA ATCTAGGCGC CTGATTGAA TCTGAAAGCG GAACAATTAT TGAAAAAGGA	3480
TAAGTTGTTT TACTAATAAG ATGTATTCTA TTTCTAGTAT CTTTATATCA AATTAGAAAT	3540
TATTCTTGAA AACATGTACA ATATTTCAAA AGATACTAGT TTTAGACTTT AATATGGTAA	3600
AACAAATATA AATAGAAAGC GTTTTCTTGA ATGTTTATTT AAGAAAGTAG TTGGTTTTTT	3660
ACACTTTGTT AGACATCAGG AGGAAAAACA AATGAGTGAA AAAGCTAAAA AAGGGTTTAA	3720
GATGCCTTCA TCTTACACCG TATTATTGAT AATCATTGCT ATTATGGCAG TGCTAACTTG	3780
GTTTATCCCT GCGGGGGCCT TTATAGAAGG TATTTACGAG ACTCAGCCTC AAAATCCACA	3840
AGGGATTTGG GATGTCCTCA TGGCACCGAT TCGGGCTATG CTAGGTACTC ATCCAGAGGA	3900
AGGTTGCTC ATTAAAGAAA CGAGCGCAGC GATTGATGTA GCCTTCTTCA TCCTTATGGT	3960
TGGTGGTTTC CTTGGCATTG TCAACAAAAC TGGTGCTCTT GACGTAGGGA TTGCCTCTAT	4020
CGTGAAGAAG TATAAGGGCC GCGAAAAAAT GTTAATTTTG GTAGTGATGC CTTTGTTTGC	4080
CCTCGGTGGT ACAACTTATG GTATGGGTGA AGAAACAATG GCCTTCTATC CACTCCTTGT	4140
GCCAGTTATG ATGGCCGTTG GTTTTGATAG CCTGACTGGT GTTGCAATTA TTTTGCTCGG	4200

559

TTCTCAAATC	GGCTGTTTGG	CATCTACTCT	GAATCCATTT	GCGACAGGTA	TTGCTTCAGC	4260
GACTGCGGGA	GTTGGTACAG	GGGACGGTAT	CGTACTTCGT	CTGATCTTCT	GGGTACCTTT	4320
GACTGCTCTT	AGTACTTGGT	TTGTTTACCG	TTATGCGGAT	AAGATTCAAA	AAGATCCGAC	4380
TAAGTCACTG	GTTTATAGTA	CTCGCAAAGA	AGATTTGAAA	CACTTTAACG	TAGAAGAATC	4440
TTCATCTGTA	GAATCTACAC	TTAGCAGCAA	ACAAAAATCA	GTCTCTTCT	TATTTGTGTT	4500
GACATTCATC	TTGATGGTAT	TGAGCTTCAT	TCCATGGACA	GACCTTGGCG	TTACCATTTT	4560
TGATGACTTT	AATACTTGGT	TGACTGGTCT	TCCAGTTATT	GGTAATATTG	TCGGTTCATC	4620
TACTTCTGCA	CTAGGTACTT	GGTACTTCCC	AGAAGGCGCA	ATGCTCTTTG	CCTTTATGGG	4680
TATCCTGATT	GGTGTATTTT	ATGGTCTTAA	AGAAGATAAG	ATTATCTCTT	CCTTCATGAA	4740
TGGTGCTGCT	GACTTGCTCA	GTGTTGCCIT	GATCGTAGCG	ATTGCTCGTG	GTATTCAAGT	4800
TATCATGAAC	GACGGTATGA	TTACCGATAC	AATCCTCAAC	TGGGGTAAAG	AAGGCTTGAG	4860
CGGTCTATCT	TCACAAGTCT	TTATCGTTGT	AAC TTATATC	TTCTATCTAC	CTATGTCATT	4920
CTTGATCCCA	TCTTCATCTG	GTCTTGCCAG	CGCAACTATG	GGTATCATGG	CTCCACTTGG	4980
AGAATTTGTA	AATGTCCGTC	CTAGCTTGAT	TATCACTGCT	TACCAATCTG	CTTCAGGTGT	5040
CTTGAACTTG	ATTGCACCAA	CATCTGGTAT	TGTGATGGGA	GCTCTTGAC	TTGGACGTAT	5100
CAACATTGGT	ACTTGGTGGA	AATTCATGGG	CAAAC TCGTA	GTCGCTATTA	TTGTAGTGAC	5160
CATCGCCCTT	CTTCTCCTTG	GAACCTTCCT	TCCATTCTTA	TAAAATAGTG	AGTGAGGTGA	5220
TTCCATGAAA	ATAGATATAA	CAAATCAAGT	TAAAGATGAA	TTTCTTATAT	CATTAAAAAC	5280
CTTGATTTC	TATCCTTCAG	TACTCAATGA	AGGAGAAAAT	GGAACACCTT	TTGGACAAGC	5340
AATCCAAGAT	GTCTTAGAAA	AAACTTTAGA	GATTTGTCGA	GACATAGGTT	TCACTACCTA	5400
TCTTGACCC	AAAGGTTATT	ACGGATATGC	AGAAATCGGT	CAGGGAGCAG	AGCTTCTGGC	5460
CATTCTCTGT	CATTTGGATG	TTGTTCCATC	AGGTGATGAA	GCAGATTGGC	AGACACCGCC	5520
ATTTGAAGCA	ACTATCAAAG	ACGGCTGGGT	ATTGCGACGT	GGTGTCCAAG	ATGATAAAGG	5580
CCCTTCGCTC	GCAGCTCTCT	ATGCAGTAAA	AAGCTTGCTG	GACCAAGGTA	TTCAGTTCAA	5640
AAAGCGCGTA	CGCTTTATCT	TTGGTACCGA	TGAGGAAACC	CTCTGGCGCT	GCATGGCACC	5700
CTACAATACC	ATCGAAGAAC	AGGCCAGTAT	GGGCTTTGCA	CCTGACTCAT	CTTTTCCTCT	5760
GACCTATGCT	GAAAAAGGGC	TTCTACAGGT	CAAAC TT CAT	GGCCCTGGAT	CGGATCAACT	5820
AGAGCTTGAA	GTAGGAGGCG	CCTTTAACGT	TGTACCAGAC	AAGGCCAACT	ACCAAGGTCT	5880
CCTCTATGAA	CAGGTTTGTA	ACGGTCTCAA	AGAAGCTGGT	TATGATTACC	AAACCACTGA	5940

560
ACAAACCGTA ACGGTTCTCG GAGTGCCAAA GCATGCTAAG GATGCTAGTC AAGGTATCAA 6000
TGCTGTCAATC CGACTAGCTA CCATTCTTGC TCCTCTCCAA GAACACCCTG CTCTCAGTTT 6060
TCTTGCAACA CAAGCAGGTC AAGACGGCAC AGGAAGACAA ATCTTTGGTG ATATAGCAGA 6120
TGAACCTTCT GGTACCTAT CCTTTAATGT CGCAGGTCCTC ATGATCAATC ATGAACGTTT 6180
TGAAATCCGT ATTGACATTC GGACTCCTGT CTTAGCTGAC AAGGAAGAAC TAGTAGAGTT 6240
GCTTACAAGA TGTGCACAAA ACTACCAACT CCGCTACGAA GAGTTTGACT ATCTAGCGCC 6300
TCTATACGTC GCAGAAGACA GTAAACTCGT TAGCACACTG ATGCAAATCT ACCAAGAAAA 6360
GACTGGCGAT AACAGTCTTG CTATTTTATC CGGTGGTGCC ACTTTTGCTC GCACCATGCC 6420
AAATTGTGTA GCCTTCGGCG CCTTATTTCC AGGAGCGAAG CAGACAGAAC ATCAGGCAAA 6480
TGAATGTGCC GTTCTAGAAG ATTTGTACCG TGCTATGGAT ATTTATGCCG AAGCCGTCTA 6540
TCGACTTGCA ACTTAATCAG GCAACTGTTT CTACCAAAAA AAATCGACCG ATTAATGAAC 6600
TGCACCCCAA AAGTTAGACA GAATAAATCT AACTTTTGGG GTGTTTTATT ATGAAATTGA 6660
GTTATGAAGA TAAAGTTCAG ATCTATGAAC TAAGAAAGCA AGGACAAAGC TTCAAACAGC 6720
TTTCAAAAAG ATTTGGTGTG GATGTTTCTG GTCTAAAGTC ATCTGAATCT TTGAGATGAG 6780
CTTTATAAAT CGCTTTTTTC AGTTTTTGCA CTGGTGTTC GATAAACTCA AACTTTTTAG 6840
CCGTGGTATT GCCTGATTTT ATAGTATATT GAACTAGAA TAGTACACCT CTCCTTCTAA 6900
AACATTTTTA GAAATCGATT TGAATGTCCT GATCGATTTG TCCTGTCTTT ATTTTCTTTT 6960
ACTATATTTG AGCCACTTCG TCTTTAACGG CTTTATTCAT AAGCTCTTGT AATTTTTCTT 7020
TACTATCAAT TACTTCTGAT TTTCCGTTGT AATTATTTGT AATAGGTTTT AACTTACCTA 7080
ATTTCTCGAC ACGCTCATT AATTGATCTT TTTTGAAGGC TGCTTATGTT TTTCTAAGA 7140
TTTTTTCAAA AATATATTTA TCAGATAGCG GTTGTCTTC TTCTTCAGCT TGGTTTTTGT 7200
ATTAATTTGA AACATAAGGA ACAAATCCTT CATAGTAACC TAATGCTCCC ATAAGTTCAA 7260
AAGCTTGTTT TCTAATTCAA ACCATTGCAA CTCAGATTTT AGCTTTTCAG ATAAATCCTG 7320
CTCATCCAAA TAATGACTTG AAATTAGTGC TGAATCGTT TCTGTATCCT GTACAGGCTG 7380
AGCACCCATA CCAGCAAAAA ATAAACTCGT TCCTAGCAAG ACCGAACAAG CTCCTATTGC 7440
ATATGGCCTC AAAGAAAAAC GCTGCTTTCT CTCAAATTGA AATTCTTTCA TCCCATTCTC 7500
CATCATTCAT TATTACTGTA TATTTTGTAT ATCAGAAATA GTTTGTATTC ACAAATCTTT 7560
CTAGTTATTC CCTTATCATT CCTAATTAAG GGAGATAACA TACAATAATT TTTAGTTAAA 7620
TGATATCGA TGTTTTTTGT TTTTCTTAAT AAACGCAATA CAAAAAGAGC CTGTTACCAA 7680
GCTCTTTGTA CTCAATGAAA ATCAAAGAGC AAATTAGGAA ACTAGCCACA GGTGCTCAA 7740

561

AACACCGTTT TGAGGTGCA GATAGAACTG ACGAAGTCAG CTCAAAACAC TGTTTTGAGG	7800
TTGCAGATAG AACTGACGAA GTCAGTAACA TCTATACGGC AAGGCGACGC TGACGTGGTT	7860
TGAAGAGATT TTCGAAGAGT ATTAGTCTAT TATTTCTTCT CAGCGCGAAG GGCTGACAAG	7920
ATTTGTGTTT GGATATCATC CACACCATTT GGAGTATTTG GTAAAAAGAT AGTTTGATTT	7980
CCTTTAGAGG CAAAGGTATT CAAGGTATCC AAATACTGGT TGGTCAAGAG GATAGACATG	8040
ATTTGTTCCT CTGTCATGCC AACATTGGCT TCCTTGAGTT CGGTGATAGA CTCTGCCAAT	8100
CCATCCACAA TCGCCTTACG TTGTTGGGCA ATCCCCACAC CATGAAGGCG GTCTTTTCT	8160
GCTTCTGCTT CAGCTGCAGT GACAATTTTA ATCTTGTCAG CTTCCGCCAA TTCTTGCTCT	8220
GCGACCCGCT TACGTTGCGC CGCATTGATT TCATTTCATGG ATTGCTTAAC TTCTGCATCT	8280
GGTTCGACCT TGGTAATCAA GGTTTTCACG ATAATGTAGC CGTAAGTGGT CATTTCTTCT	8340
GCTACTTGGT GTTGAACATC AAGGGCAATC TCATCTTTTT TCTCAAACAA TTCATCCAAG	8400
GTTAATTTTG GAACAGAAGA GCGAAGAGCA TCTTCGATAT AAGATTTAAT CTGAGATTCT	8460
GGACGTATGA GTTTATAGTA AGCATCTGTC ACGCTCTGCT CGTTGACACG GTACTGAGTC	8520
GCTACATTCA TCATAACGAA CACATTGTCC TTGGTCTTAG TCTCAACCAC AATATCACTT	8580
TGCAACAAGC GCAACTGAAT CCGTGCTGCA ATCGAGTCAA TCCCAAAAGG CAAGCGAATA	8640
TGAATACCGC TATTAGCAAC CTTTGGGTAT TTCCCAAAGC GTTCAATAAT CGCCACCGAC	8700
TGCTGACGAA CCACATAAAC TGTACTCAGT GTGACTATCA CCAATAGGAG CACACAAACA	8760
ATCAGAAAAA TCATGAAAAA TATTGCCATA ATGGAACCTC CACAAGTATT TTTCTAGTAT	8820
TATAGCACAT TTAAAGAAGG CTGTGCCGTT TTTACTGCGA TTTTTCCTGA AATGTCAATA	8880
ATTAGAGGTG AATTGTCCTA TTGTCGTCCA ATCTCTTGCT AAAATAACTC TTTATAAAAG	8940
GCAATCGTTT CTTCTAAGGT TGGCATAAAT GGATTCCTG GTGCGCAGGC ATCAATCAAG	9000
GCATTCTTAG AAAGGTATTC AAAGTCGAAA TCTTTTTCTT CAATACCAAG TTCAGTCAGT	9060
TTCTTAGGAA TACCTACTGT CTCAGAAAGC TTCTCAATCT CAGCAATCGC ATAATCGGCA	9120
CATTCTTGAT CTGATTTACC TTCTACATGA AGTCCCAAGG CTTTGGCAAC ATTGCGGAAA	9180
GCTTCTGGTA CACGTTTAGC ATTTTCACGT TCTATAACTG GTAGCAACAT GGCACAGCAC	9240
ACGG	9244

(2) INFORMATION FOR SEQ ID NO: 69:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8898 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double

562

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

GATCTGAACT TTATCATCAT AACTTAATTT CATAATAAAA ACACCCCAA AGTTAGATTT	60
TTTCTGTCTA ACTTTTGGGG TGTAAGTCAG TCATTGGACT GACGTTTTTT TGTATGCTTA	120
TTTTGATTG ATGTAGTTGA TACCATCTGC TTTTGGTGCG ACTGCTTTTC CAAAGAAGGC	180
TGCTAAGACA AGAATTGTCA AAACATAAGG TGCAATTTGA AGATAAACCG CTGGCACTCC	240
TTGTAGGAAC GGCAATGAG AACCGATAAC AGCCAAACTT TGTGAAAGTC CAAAGAAGAG	300
ACTAGAAAGC ATAGCACCGA TTGGATTCCA TTTCCCAAAG ATCATCGCAG CAAGGGCGAT	360
AAATCCAGGT CCAACAATAG TTGTCACTGA GAAGTTAACT GAGATTGATT GCGCATAAAT	420
CGCTCCGCCA ATTCCACCTA GAAAACCTGA AATAATAACC CCTAAATATC TCATCTTGTA	480
GACGTTGATT CCCAAGGTAT CCGCTGCTTG AGGATGTTCA CCGACAGAGC GGAGACGAAG	540
ACCAAATTGA GTCTTAAAGA GAATAAACCA AGCAAGGAAT GAGAAGGCAA TCGCCAGATA	600
ACCAAGTAGA CTAGTTGACT TGAAGAAGAT ATCACCAATC ACTGGGATAT TTGCCAAGAC	660
TGGGAAATCA AAGCGTCCAA AAGTTGACT TAGGTTGTCG GTTGTCTCTT TGTATAAAG	720
AACTTTAACT AAGAAAACAG CCAAGGCAGG CGCCATCAAG TTCAATACCG TACCGCTGAC	780
AACATGGTCT GCACGGAAT GAACCGTCGC TGCTGCGTGG ATGATAGAGA AAACACTACC	840
AACCAATCCT GCTACAAGCA AGGATAGCCA TGGAGTTGCT GCTCCAAATT GTTCTGCAAA	900
TTCAAGGTTA AAGACAATC CAGAAAAGGC ACCCATAACC ATAATTCCTT CAAGGCCAAC	960
GTTTACCACA CCACCACGTT CAGAGAAAAC ACCACCGATA CTTGTAAAGA TGAGAGGTGC	1020
TGAGTAAATC AGCATAGAAG ACACCAAGAG GGGGAGCAAG GTTATAATAG ACATCTTTAC	1080
TTACCTCCTT TAACTTGTTT TTTCGGTTTG ACAAAGCGTT CGATAAGGTA ATGAACACTG	1140
ACAAAGAAGA TAATAGACGC TGTTACAATG CTGACAAGCT CAGATGGTAC CTGCGCCGCA	1200
TTCATACCAG GAGCCCCAAC TTGGAGAACG CCAAATAGGA AGGCTGCAAA GAGTATACCA	1260
ATTGGTGAGT TGGCCGCAAG CAAACTAACC GCCATTCCGT TAAATCCGAT AGCTAATGAC	1320
GAACCTTGAA CATAGACGTT CTGGAAGGTT CCCAAACCTT CAACAGCTCC ACCAAGACCT	1380
GCCAAGGCAC CTGAAATAAT CATAGATAGG ATAATAGTCC GCTTGGCAGA AATACCAGCA	1440
TATCTGAAG CATGTGGATT AAGACCAACT GCACGGATTT CAAAACCAAG AGTTGTTTTTC	1500
TTGAGCATGA ACCAAATAAC TGCAACGGCA ATGATGGCAA AGAAAATACC AATATTCATC	1560
CGTGAGTTAC CAGTCAACTC AGCCAACCAA GGTGTCTGAT AGGTTGCATT AGCCCCAACA	1620

563

CGAATGGTCG AATCTGTACT TTGCATGAAG TCTTTAGGGA AAGCATGGAT AAAGGCATTC	1680
CCTACATACA AGACAATGTA GTTCATCATG ATGGTTACAA TAACCTCTGA CGTCCCTAGA	1740
TAGGCCCTAA GAATACCTGG AATCGCTCCG ACAATCCCAC CAGCAATCAA GGCAATCACG	1800
ATGGTTGCTA GAATCATCAA GGGACGGGGC ATATCTGGAT GCGACAGGGC AAACCAACCA	1860
CTGAGAATCC AACCTGCCAA AGCCTGACCA GGAAGTCCGA CGTTAAAGAA ACCAGCTCGA	1920
CTGGCAACGG CAAAACCAAG ACCAATCAAG ACCAGAGGAC CCATAGCACG GAAGATTTCT	1980
CCAATCCCAC GCAGACTGCC AAAGGCTGTA TAGAACAATT CTTCGTAGCC CCAAATAGCA	2040
TCATAACCGA AGATCCACAT GACAAATGGCT CCGAGTAAAA TTCCTAGGAA TACAGAAATC	2100
AAGGGAACCG AAATTTGTTG TAATTTTSTA GACATCACTC TTCTCCTTTC CCAAGTTTCC	2160
ACCAGCCATC AAGACACCAA GTTCTGTGTT ATGGGTGTT TCTGGTGATA CAATACCTTG	2220
AATCTTACCA TCGTGGATAA CGGCAATACG GTCTGAGACG TTTAAATCT CATCCAATTC	2280
AAAGCTGACA ACAAGGACAG CCTTGCCATT ATCACGCTCT TCAATCAAGC GTTTGTGGAT	2340
ATACTCAATG GCACCGACAT CCAACCCACG AGTTGGCTGG CTAACGATAA GGAGATCAGG	2400
ATCTCGATCA ATTTCACGAG CAATAATTGC TTTTGTGTA TTTCCTCCTG AGAGTGCAGC	2460
TGCAGGAACT AATTCACTGG CAGCGCGAAC ATCAAACCTCT TCCATCAGCT TTTTAGCATA	2520
AGAAGTAATA TTTGAATAAT TCAAAATTCC ATTTTACTA TGTGGTTCTT TATAGTAGGT	2580
TTGAAGGGCA ATATTTTCAG ATATCATCAT TTCCAAAATC AAGCCATCAC GGTGACGGTC	2640
TTCTGGAACG TGCCCAACAC TTAGTTCTGT AATCTGACGT GGGTGCAAGC CTACAATTGA	2700
ATCTCCTTTT AGCTCAATGC TACCAGATTC AACCTTACGA AGACCTGTAA TGGCTTGAAT	2760
CAGTTCAGAC TGACCATTTT CATCAATCCC CGCAATACCA ACAATCTCTC CAGCACGAAC	2820
ATCCAAGGAC AGATTTTSTA CAGCTGGAAC ACCACGGTTT TCATTGACCA CCAAATCTTT	2880
GATAGACAAA ACCACTTCTT TTGGTTTAGA GGCTTGCTTC TCTGTTTTAA AGGAAACAGA	2940
ACGTCTACC ATCATTTCCG CCAATCAGC ATTGGTAGCC CCTGCAATTT CAACGGTTTC	3000
AATTGATTTT CCACGACGGA TAACTGTAAC ACGGTCAGAA ACTGCTCGAA TTTCATCCAA	3060
TTTGTGGGTA ATCAAGATAA TTGATTTTCC TTCTTTGACA AGATTTTSTA TAATAGCCAT	3120
CAACTCATCA ATTTCTGATG GAGTCAAAAC AGCCGTTGGT TCGTCAAAGA TAAGGATATC	3180
AGCCCCCGA TAAAGTGTTT TTAATTTTTC TACACGTTGT TGGGCTCCAA CTGAGATATC	3240
TGCTACCTTG GCAGAAGGGT CAACAGCTAA GCCATAACGT TCAGAAAGAG CCTTGATTTT	3300
TTTGCTAGCT CCAGCGATAT CTAGCACACC ATTTTATGTC AATTCATAC CTAAATGAT	3360

564

GTTCAGCC ACTGTGAAGG CTTCAACCAA CATAAAGTGC TGGTGAACCA TCCCGATTCC	3420
CAAGCTAGCT GCTTTAGATG GGGAGTCGAG ATTGACAACCT TGACCGTTGA CCGCGATTTC	3480
ACCACTAGTT GGTTCAGAA GGCCTGCTAA CATGTTCACT AGCGTGGACT TACCAGCCCC	3540
ATTTCTCCT AAAAGTGCAT GAATTTACCC TTTTCGTAGG TGCAAGTTGA TTTTGTCTGT	3600
GGCAACAAAT CCACCAACA CCTTGGTAAT ATCAGCATC TCAATGACAT TTTCTGTGC	3660
CATGTGCTCT TCCTTCAGA GTCTTATTTT ATTTCAATAA AACTTGCTAG TTTGTCTAGT	3720
AGCAAGCTTT ACTTAGACAA AATGACTTTG TCTCAACTCT TAAAAAGCG GCCCTTGGCC	3780
GCTTCCTAAG AAATGACTTC CATCCATTAT TTTTCAGGAA CTTTACGCT TCCATCAAGG	3840
ATTTTAGCTT TTGCATCTTC GACAGCTTTT TTACTTCTT CTGAAAGGTT TGTACTGCC	3900
AAGTCAACCC CTTTATCCTT CAATGAGTAA ACGATCACTT GACCGCCAGG GAATTCCTCT	3960
CTTCTGCCT TGTAGAAAT ATCTTTTACA GTGTACCAA CTGTTTCAA AGTAGATACA	4020
AGAACAAAGT TTGATCTTTT GCCATCTTTA GAAGTGATTT TACCTTCTGC TTCTTGGTCA	4080
CGATCAACAC CGATAACCCA AACTTTTCA TTTTCAGGAC GGCTTTCGTT GAGAGATTTT	4140
GCCTCTGCAA AGACACCTGC ACCTGTACCA CCAGCTACTT GGTAAACAAT ATCTGCACCG	4200
GCTGCGTATT GTGCGGCTGC AATTGTTTTA CCTTAGCCG CATCACCAA TGAACAGCG	4260
TAGTCAACTT GGACTTTGAT AGATGGGTCT ACTGACGCAA CACCAGCCTT GAATCCTGCT	4320
TCAAAACGAG AGATAACTTC AGATTCGATA CCACCTACAA AACCAACTTG TTTTGTCTTA	4380
GTTGTTTTTG CTGCAGCCAC ACCTGCAAGG TAACCTGACT CATTATCAGC GAAAGTTACG	4440
CTCGCAACAT TCTTTTGGTC TTTAATCACA TCATCAATCA AGACATAGTT CAAGTCAGTG	4500
TGTTCTTTTG CTGCATCTTT AACTGCATTA TTAAGGGCAA AACCAACACC GAAGATTAGG	4560
TTGTAACCTC CAGCCGCTTG TTGCAAGTTG TTAGCGTAGT CAGCTTCACT TGTTGATTGG	4620
AAGTAAGTGA AACCGTTATC TTTTGAAAGA TTGTGTTCTT TACCCCAAGC CTGCAACCT	4680
TCCCAAGCTG ATTGGTTGAA TGATTGTCA TCAACACCAC CAGTATCAGT GACGATTGCT	4740
GCTTTTGTCT TCACATCAGA AGATGAAGCT GCGTTACGAG AAGAGCGGTT ACCACATGCA	4800
GCAAGTCCAA CTGCTGCCAC TGCAACTAGG CCAAGACCTA GCCATTGTTT CTGTTCATT	4860
ACTGAACCTC CTAAATAAGA TGTGCAACGA TGTGCAAGT ATGGATTGGT TGGCCACAAG	4920
GACCGTGCCA CTCAGAGAGC GACTCAGACT AGTTAAGTC TGTAAGAGAG TATGGAAGTA	4980
ATTCCCCGAC CGTCATCTCG ACCGTCGATT TATCTTTTGC GACTAAGGTC ACTTTTAGAT	5040
CTTGTTCAAA AAATTCAGCC ATCACTTGGC GACAAGCACC ACATGGCGAG ATCGGTTTTT	5100
CAGTTTGACC ATAGACAATC AATTCTGAAA ATTCTCTTTG GCCTTCAGAT ATAGCCTTAA	5160

565

AAATAGCTGT TCTCTCACCG CAATTGGTCA AAGGATAGCT AGCATTTTCA ATATTTCACTC 5220
CCGTGTAAAC ACTTCCGTCT TTAGCTACTA AAAGTCTCC GATAGGAAAG TGAGAATAGG 5280
GGACATAGGC ATGTTTGCTG GTTTCATTTG CCAGTTCAAT CAACTCAGTA GTCGCCATCT 5340
GCCAATTCTC CTTTAAAT AGCTACCCCA GCTGACGTTT CGATACGGGT CGCACCTGCT 5400
TCGACAAAGG CAAGAGCATC TGCATAAGAA CGAGCTCCAC CGGCGGCTT GACACCCATA 5460
TCAGATCCAA CTGTTTCACG CATTAAATGA ACATCTGCTA TCGTAGCACC ACCAGTTGAA 5520
AAGCCAGTAG ATGTTTTCAC AAAGTCAGCC CCAGCTTTT GGGCCAATTG GCAAACAACA 5580
ACTTTTCTT GGTCTGTCAG AAGGCAAGCT TCAATAATGA CTTTCACTAA CTTATCACCA 5640
CTTGCTTCCA CTACTGCGCG AATATCTGAC TCAACCAAGG CTAAATTACC TGATTGAGA 5700
GCTCCAACAT TGATCACCAT ATCAATCTCA TCTGCACCAT TTTGGATAGC TTCTTTTGTC 5760
TCAAATGCTT TCACGGCTGA AGTTGTTGCT CCCAAAGGGA AACCTACTAC TGTGCAAACC 5820
TTAATCTG TGCCTTCAAG TCCTTTTGA GCATGTTCAA CCCAGGTCGG ATTAACGCAA 5880
ACACTGGCAA AGTCATACTC TCTAGCCTCA GACAACAAAC TATCAATTTG TTTTCTTTT 5940
GCATCTTGT TTAAGCGT ATGATCTATA TATTTATTTA ATTTCAATTC GGTTTTCCCT 6000
CCATTTAGGA GATGATTTCT ACAATTTTAC GGATTTTTC CACTTCATCA CTTATTTTAA 6060
CACATTTTG GAAATCTGTA ACTAGTTGAG GTGGAATTTT TTCATTTGTG TATACTTTTG 6120
CAACAATTC ACCCTTTTGA ACGGAGTCTC CAATCTTCTT TTCAAAAACA ATTCCTGTTT 6180
CATAGTCCAA GGCATCAGAC TTAAGTCAC GACCAGCACC CAGCCTCATG GCATAAAGAC 6240
CAAAGTCCAT AGCTGGAAGA GCTGAAATGA CACCCGTTTC CTGAGCAGG ATTTCCACCA 6300
CATGAGCTAC ATTTACAGGA CGATAGAGGT CTCCAAGTC TCCACCTTGG GCTTGACCA 6360
TTTCCTCAA CTTAGCCAGT GCTTGACCAT TCTCAAGATG TTGGTGAAT TCTTCAACAG 6420
TTTTGTAAAC ATTTGCCAAA CCAAGCATAA TTTGAGCCAA TTCACAAATA AAGTGGGTAA 6480
TATCCTGACG TCCTTGACCT TGCAAAATCT CCAATGCTTC AAGGATTTCC AGACGATTTT 6540
CAATCGCTCG TCCCAAAGGC TGGCTCATAT CCGTAATCAC TGCTACTGTC TTCCGTCCAA 6600
CAACCTTACC AAGATCTACC ATAGTTTGAG CCAACTCAG CGCCTCATCA ACCGTCTTCA 6660
TGAAGGCACC CTCACCGACA GTCACGCTA GCAAAATAGC ATCCGCCCCC GCCGCAATTT 6720
TCTTGCTCAT CACCGAACTC GCAATCAAAG GAATCGTGTC GACAGTTGCG GTCACATCAC 6780
GAAGGGCATA GAGAAGCTTA TCTGCTTTGA CCAGCTGGTC TGATTGCCCA ATGACAGATA 6840
CTCCAATATC CTGAACCTGA CGAATAAAT CCTCTTGACT ACGTTCTACT TGATAGCCCT 6900

566

TAATGGACTC CAATTTATCA ATTGTTCCGC CTGTATGGCC AAGACCACGA CCACTCATTT	6960
TTGCTACAGG CACACCGAAG CTAGCAACAA GAGGAGCTAA AATCAAGGTT ACCTTATCGC	7020
CGACACCACC AGTAGAATGC TTGTCAACTT TCACACCATC AATGGCTGAC AGGTCAAAC	7080
CTTGCCCAGT CTTAACCATA TTCATCGTTA AATCAGAGAT TTCTCGAGTC GTCATTCCCT	7140
TAAATAAAC AGCCATAGCA AAGGCAGACA TCTGATAATC AGGAACAGTT CCTGATACAT	7200
AGCCTTCTAT CAGCCATTCA ATTTCACTTG AAGTCAGTTC TTGACCGTCT CGTTTTTTTT	7260
GGATTAAATC AACTGCTCTC ATTCTTTCAC ACTTCTAAGG ATATAGTATC CCTTGTCTTT	7320
TTTAAGGATT TCACAATGTC CAAACACATC TTCCATCTTA GACTTGGCAC TTGGAGCTCC	7380
TTGTTTTTTC TGGATGACGA TGGTCAAATC TCCACCAATT TCCAAGAAAT CTTTACTTTT	7440
CTCGATGATT TCATGAACGA CTTGCTTGCC CGCACGGATA GGAGGATTGG AAATGACATG	7500
GTCAAATCGC CCTTGAATC TTGCATAAAT ATTAGATTGA AATATCGTCG CTTTTGCATT	7560
ATTTTTTTCA GCATTTCTCT GAGCTAAATC CAGGGCACGA GTGTTAATAT CAACCATGGT	7620
CGCCTGAAC TCGTAAACCT TGACCAAGGA CAAACCTAAT GGACCATAAC CACAGCTAC	7680
ATCTAGGACT GTCTCTCCTT GGTGACATC CAGACACTTG AGCAAGAGTT GACTTCCAAA	7740
GTCAACCAT TTTCTGCTAA AAACCCCGC ATCTGTCAAA AAAGTCATTT TTTCTCCAA	7800
CAAGTCCACT CTCAACTCAT GAATGTCGTG AGCAGCGTCA GGATTTTCTG CATAGTACAT	7860
TTTACTCATG ACACTATTTT ACCATAATTT GACTCAAATT GTAAATCGTT TACAAATTGA	7920
TAATAAAACG AAAAAGACCG AAGAAAGCAA GTCACGAAGC CATTTTCTTC AATCTCTTTC	7980
AACACTTATA AATAATAAAC CATTTAGAAC TATAAATATC AACTCCAGA TAAAAACAA	8040
AAGTTTATCA TCTATAATCA GGCAGATTAT TATTTCTATT GCTTAACCTT AAAATACTTT	8100
ATTATCAACA AAATTCCTAA CAAATGTTT AGATAAAAGC CCAACTGATA CGTTTATGTC	8160
AGGATTTCCA AACTTGTCCA AAGTCGTATC AAATCTTCTA GTGACATGTG GAAGAAATAA	8220
CCCTCTGTG CAATCCGTAG GACTAAAAAG CAATAACTAC CCGCAGCAAT CCATTTCGTC	8280
CATCGTTTTT TAGTAAGAAA GCAATTAAGA ACGAACAAT AAAGACAGCT GTTACAATAG	8340
CATGTTCCAT CAAAAAGTA AAACCGTAAT AGGTTTCCAC AAAGCATCTA CCATTATCTG	8400
CATGGTTTCC TTTTATAAAA GGTAAAGCAA AACTTAAAT AAAACAGAGT TCCAATATGT	8460
AACGTTTTAA GATTTTCATA GTACACCTCC TATAAGTTGT GAACTAAAAA GCCCCCTTTA	8520
TAAGCTTATA AATCAGTAGA ATCTATCTCC TATTTTATCA ATAAATTGAT CACTTATACT	8580
ATATACCATT GACTTACCAC ATTCAAGAAA CCGCTTTATT TTTTAGCTT TTTATGGTAT	8640
GATAGACAAA ATATCTAGGG GAAAACAAAT GACCAACGAA TTTTACATT TTGAAAAAAT	8700

567

CAGCCGCCAG ACTTGGCAAT CTTTACATCG AAAGACAACA CCTCCTTTGA CAGAAGAAGA	8760
ATTGGAATCT ATCAAGAGTT TTAATGACCA AATCAGTCTC CAAGACGTTA CAGATATCTA	8820
TCTCCCCTTG GCTCATTTGA TTCAGATTTA CAAGCGAACT AAGGAAGATT TAGCCTTTTC	8880
AAAAGGAATT TTCCTCCA	8898

(2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13188 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

TATCTTAACG aGGATTGGGT TTATCGTCAG TCTTATTGCC CTAATTGTGG GAACAATCCC	60
TTAAATCATT TTGAAAATAA TCGGCCTGTA GCAGATTTTT ACTGTAATCA TTGTAGTGAG	120
GAGTTTGAAC TAAAGAGCAA AAAAGGAAAT TTTTCATCAA CAATCAATGA TGGTGCTTAT	180
GCAACGATGA TGAAGCGTGT GCAGGCAGAT AATAATCCTA ATTTCTTTTT TTAACTTAC	240
ACAAAAAATT TTGAGGTAAA TAACCTTCTT GTCCTTCCGA AGCAATTGTGT TACACCGAAA	300
TCGATTATTC AAAGAAAACC ACTTGCACCA ACTGCTAGAC GAGCAGGTTG GATTGGTTGT	360
AACATTGATT TATCACAAGT ACCTTCTAAA GGAAGGATAT TTCTTGTGCA AGATGGACAA	420
GTTAGAGATC CAGAAAAAGT TACAAAAGAA TTTAAGCAAG GTTTATTTTT AAGGAAGAGC	480
TCTCTGTCAT CAAGAGGTTG GACAATAGAA ATTCTAAATT GTATAGATAA GATAGAGGGT	540
TCAGAATTTA CCCTTGAAGA TATGTATCGT TTTGAAAGTG ACCTAAAAAA TATCTTTGTT	600
AAGAACAATC ATATCAAAGA AAAGATTAGG CAACAGCTTC AAATATTAAG AGACAAAGAA	660
ATAATAGAAT TTAAAGGTAG AGGAAAGTAT CGGAAATTAT GAAAACGAAA CAACTTGTGT	720
CATCAGAAGA GGTGTATGAT TTCTTAAAAG TCATCTGGCC TGATTATGAA ACTGAAAGCC	780
GTTACGATAA CCTAAGTTTA ATCGTCTGTA CCTTATCAGA TCCCAGTTGT GTGAGATGGT	840
TATCTGAAAA TATGAAATTT GGTGACGAAA AACAACTAGC TTTGATGAAG GAAAAATATG	900
GGTGGGAAGT AGGAGATAAA TTGCCAGAGT GGCTACATAG CTCCTATCAT AGATTATTGT	960
TAATAGGTGA ATTATTGGAA AGCAATCTAA AACTGAAAAA GTATACAGTA GAAATTACAG	1020
AAACTTTATC ACGTTTAGTA AGTATAGAGG CTGAAAATCC AGATGAAGCC GAACGACTTG	1080
TAAGAGAAAA GTATAAGAGT TGTGAAATTG TTCTTGATGC AGATGATTTT CAGGACTATG	1140

568

ACACTAGCAT ATATGAATAG GTAGATGTTT TTATTTTGTC AACAAAAAG AGGCTCGCAC	1200
CTCTTTTCT TATTCTTTT TATGATTAA TACGGCATTG AGGACAATAG CGAGTAGGCT	1260
GGCTACGACG ATTCCGTTTG AGAAGAACAT TTGGAAGGCT GTCGGCATGC TGACAAAGAG	1320
ATTACTGTTG TTGAGACCGA CACCTGCAGC GATTGAAACA GCTGCGATAA GGAAGTTGTG	1380
TTCATTGTTA GCAAAGTCAA CACGGGCGAG GATTTGATC CCTTGAATTG ATACAAAACC	1440
AAACATTACC AGCATGGCAC CACCGAGGAC GGAGCTTGGG ATGATTTGGG CAAGGGCGCC	1500
AAACTTAGGA AGCAGTCCAA GGAGAACCAG GAAACCAGCT GCGTAGTAGA TTGGCAGGCG	1560
TTTTTTGATG CCTGACAATT TAACCAAACC AACGTTTGT GAAATCCGG TGTAAGGGAA	1620
GGTGTAAAG ATTCTCCGA GAAGTACGGC CAAACCTTCT GCGCGGTATC CGTTGCGAAG	1680
GCGCGTGCTG TCGATTGGAT CCTTTGTGAT ATCAGACAAG GCCAGATAAA CACCAGTTGA	1740
CTCAACCATA GACACCGTTG CGATGATACA CATCATGACA ATAGATGAGA TTTCAAAGGT	1800
TGGCATCCCA AAGTAGAGTG GAGTTGGGAC ATGGACAAGT GGAGCTACCG CAACAGGAGA	1860
GAAGTCCACC AAGCCCATAG TAGCAGCAAT GGCAGTTCCA ACAACCAGAC CAATCAAAT	1920
AGAGATAGAC TTGATAAATC CTTTGGTAA GATGTTGATC AAGAGGATAA TCAGAACAGT	1980
AATAGCTGCA AGCAAGAGAC TTTGACCAGT TGGCTCTGGA ACGTTATTC CATATTTCC	2040
AATAGCGACA GGGATCAAGG TTAACCAAT CGTGGTAATA ACAGATCCTG TTACGATAGA	2100
TGGGAAGAGA TTGGCTACTT TTGAGAAGAT GCCTGAAACA AGAACCACGT AAATCCCAGA	2160
TGCGATAAGG GCACCAACA TAGCGCCACT ACCATGGCTT TGCCCAATCA TAATCAAGGG	2220
AGCGACCGAC TGGAAATGCA CTCCAAGAAC GACTGGGAGT CCAATCCCAA AGTATTGTT	2280
GAGTTGGAGT TGGAGGAAGG TTGCCACCCC ACACATGAAG ATATCTGTAG AAATCAGGTA	2340
GGTCAACTGC TCAGCTGAAT AGCCAAGGGC TGTCGCAATC ATGATGGGAA CCAGGATAGA	2400
TCCTGAGTAC ATGGCTAGTA AGTGCTGCAA GCCAAGAACG GCTGCTTGCG AGTGTTTTTC	2460
TTGAGTTTGC ATTAGAGATC TGCCTCCTTA AATACGACTT GACCATTTTC AAAACAATCC	2520
AAACGAGCAA GTGATAGGAC AGGGTAGCCT GCTTTTCAA GCAAATCACG ACCATCTTGG	2580
AAGGATTTCT CAATCACCAT ACCGATAGCT TGGACTGTGG CACCGGCCTG TTCGATGATT	2640
TGAATCAAGC CTTTAGCAGC TTGGCCATTA GCAAGGAAAT CGTCGATAAT CAAAACCTTG	2700
TCCTCTGGTG AGAGGAATTT TTCAGCGATA GAAACGGTGC TGGTCACCTG CTTGGTAAAG	2760
GAGTAGACTT GAGCAGTTAA GATGCCTTCG TTCATGTTGA TGTTCTTAGC TTTTGTGGCG	2820
AAAATCATGG GAACGTTTAA GGCTTCAGCT GTAAAAACGG CTGGGGCAAT ACCCGACGCT	2880
TCAATGGTTA CGACCTTGGT AATGCCAGTA GTAGCAAATT TTTCCGAAA AACCTTACCA	2940

569

ATCTCTCGCA TCAAGCTAAA GTCAACTTGG TGGGTTAAAA AGGAATCTAC CTTGAGGATG	3000
TTATCACCCA AGATATGCCC ATCCTTGAGG ATGCGCTCTT CTAATAATTT CATAAGACCT	3060
CCTAAAGTCT AAAAGTTAAT TTACTTGTTG TTTAAATATT TCTATAGTGA TCCCTTTTGC	3120
TAATACTATA TATTTGATAA AACTATTACG AGCGAAGCGA GTCTTATCAA ATATTTCCCG	3180
TTGTAGTGGT ATCATAGACA ATAATCTTGT TATTGTCTAT GACGGGATTT TTGAGAGTAA	3240
AATAGTTCGG GGAACATTTT TAGCCTAAGC CTAGAAATGA AAGAGCTAGG GGCTCAAAAA	3300
TTAGGGATGA AATTCCTTGG ATTCCTGAAA TTATTCACAG GATAATTTCA CCTCCCGTCC	3360
GCACTAATTA AGGGAAATAT TAAAAAAGA CCTACTTAAT CTCTAAGTAA GTCCCTTAAA	3420
TAGACATGGC AAAAACGGCC ATATCTCACT GCTGACTTAC TTATTGTTAG GTGTTCCGGC	3480
ACCTTGTAGA AACGTCGTGC CAATTCACGA CATAACAAG TAAAACGATA TTCAATTTTA	3540
AATAGGCTTG AGCCAATGTT TTTATTTTAC ACTAAATAAC TTTAGAAATC AACTATTTTG	3600
TTAGTGTTTT GGTTTAAAAA ACGAACAAAA AGAAGAGAGG GTGAACAAAA ACTCCATGTG	3660
AAGCTAACAG TTATACTAAA TGAAAATCAA AGAGCAAACCT AGGAAGCTAT CCACAACCTC	3720
AAAACACTGT TTTGAGGTTG TGGATAGAAT TGACAGAGCC AGTATCATAT ACCTACGGTA	3780
AGGCGACGTT GACGTGGCTT GAAGAGATTT TCGAAGAGTA TTAGAAGATT TTTCCATCAT	3840
AAAAGGCATA CTATCAAGCT TTTAGACACC TGACAATATG CCTTTTCTTA ACTTTAAAGA	3900
CTTTTCCCAA TTTTATTAT TCTACTCGCT AAATCTTAAA AAATAGCCAT CTGGATCCAA	3960
AACTGCAAAT TTATGAGGAT AGATATAGGG ATCACTGACA CGAAACTTTC TTTTGGTCAA	4020
GGGACGATAA ATAGGATAGT TTGCCTTCAT CACTCTTTAA TAGAGTTTGT AACATCCTT	4080
TATGCCAAAG GAGAGATTGA CTCCACGACC AAAGGGATAG GTCAGTTCAG CTAGTTGATC	4140
CTTTGTTCCT TCCTCTAACA TTAGTTGACA CTCTTCAAGA GAAAGAGAAA GTTTTCTTCT	4200
GGACGTTGGT ATTCAATCCT AAAACCCAGT AAACCACAGT AGAAGGACCG GGACTGTTCTG	4260
ATATTCGATA CAAGCAATC GGGAAATGACC GCATTGTAGT CCATATAGAA AATCCTTACA	4320
AGTCAATTTT CAAGACAATC GGTGTATGGT CTGGCGAGC ACCTGAGTCA ATCATATCAG	4380
ATTTAGTGAC CTTGTCAGCG ATACGTTTAC TTGTGAGCCA GTAGTCGATT CTCCAGCCTG	4440
TATTGTTGAT TTTAGAAGTT TTGCTGCGTT GTGCCCACCA AGTGTAGCGT TCAGGAACAT	4500
CGCCATGAAC ATGGCGGAAG GTGTCTGTAA ATCCAGTTGC CAAAAGGTTG GTAAATCCAG	4560
CACGTTCCCTC GTCAGTAAAT CCAGGTGAAC GCGGTTGCT AGCAGGATTT GCAAGGTCGA	4620
TTTCATTGTG GGCTACGTTG TAGTCACCGG TCGCAAGGAC TGGTTTTTCT TTGTCTAGTT	4680

570

CAGCCAAATA CTCAGCATAT TTGGCATCCC AGACTTGGCG TTCTTCCAAG CGTTTGAGAC	4740
CGTCACCAGC GTTTGGAGTG TAAACTTGGG TTACGAAAAA TGCATCAAAT TCTAGAGTGA	4800
TGATACGACC TTCCAAGTCC ATGGTAGAAG GGCACCGAT TTCTGGGAAG CTGATAGTAG	4860
GTGTAAGTTC TTTCTTATAA AGGAACATGG TTCCAGCATA GCCTTTACGG GCAGGCTCTT	4920
GGGAAGAGCG CCACGTGTTT TCGTAGCCTG GGAAGAGTTC TTCTAAAATT TCCACGTGTT	4980
TCTTTGTAGG TCCTTTGGCA GAAAGCTTGG TTTCTTGGAT AGCAATGATA TCAGCATTTT	5040
CAGCGACCAA GGTTTGTAGG ACTTCTTGGG ACAATTTGGC ACGAGCTGAG TCACTAGTTA	5100
GGGCAGCGTT TAGGGAATCA ATATTCATG AGATAAGTTT CATAAAGTTA CCTTTTTCAT	5160
TCAGATTATA GATTTTATTA TACCAAAAAA AGATCTATTT CCCCAACGTA TGGTTTGAAA	5220
AATTACTCTC TTTCGTTTAT AATTAAGAAT GATTTTATGA AAGGGAGTGA AAATACATGA	5280
AATCTACTC TTATGACTAT GTACTCAGCC AAATCGGTCA GCAAAATGGT ATCATGGTTG	5340
GCTTTGGGAT TGTTCATTA GCTGTGACAG TTTTTTTTGC TTTCAAGGCA TACCATAATA	5400
AAAAGGGAAG CGAATTCGT GAGTTGGTCA TGATTTCAGA TCTGGCCTTA TTTAGCTCTG	5460
CTTTTGGTCA GCATCAGGAC TTATCAAAAC AATCAAGTTT CTAACAATAA ATTTCAAAC	5520
TCACTTCATT TCATCGAGGT TGTTCCTCAA GATTTGTGAG TAGACAAGTC AGAAGTCTAT	5580
GTAAATACTT CCACAAACAC AGATGGCGCA CTTATCAAGG TGGGAGATCG CTATTATCGT	5640
GCCCTAAATG GAAGTGAGCC AGACAAGTAC CTGTTAGAGA AAGTCGAATT GTATAAGACA	5700
GACGCAATG AACTGGTGGA TGTGAACAAA TGACACTTAA TTATATCGAA ATTTTAATCA	5760
AACTGGTCTT GACTCTCAA TAGCTCAACA ACAATGTTCA CTTTGTGAAA CGTTTGATTG	5820
ATGGTAAGCC AACTCTCCTT ATCAAAAATG GGAATATTGA CCCAGAAGCC TGTCGTTTCA	5880
TTGGTTTGTG TGCATCGGAT GTATCCCTCA AACTTCGTAG CCAAGGGATT TTCCAGATGA	5940
AGCAAGTCAA ACGAGCTGTG CAAGAGCAAA ATGGGCAACT CATCGTTGTG CAAATGGGAG	6000
ATGAAAATCC TAAGTATCCA GTTGTGACTG ACGGTGTGAT TCAAGTAGAT GTCTTGGAAT	6060
CGATTGGTCG TAGCGAAGAG TGGTTGCTTG ATAACCTCAG TAAACAAGGG CATGACAATG	6120
TAGCCAATAT CTTTATTGCT GAATATGACA AGGCTGCTGT TACAGTCGTA ACTTATGAAT	6180
AAGAAAAACC TGGGGTCTTG TACTCTTCGA AAATCTCTTC AAACCGCGTC AACGTCGCCT	6240
TGCCGTATGT AGGTTACTGA CTTCTGTCAGT TCTATCTACA ACCTCAAAGC AGTGCTTTGA	6300
GCAGCCTGCG GCTAGTTTCC TAGTTTGCTC TTTGATTTTC ATTGAGTATT GGCCTCAGGT	6360
TTCCATTGTC AATCAGAAAG GGATTTTATG TCCATTATTC AAAAATTG GTGGTTTTC	6420
AAGTTAGAAA AACGCCGTTA TCTAGTCGGA ATTGTGGCCC TGATCTTGGT TTCCGCTCTC	6480